

Introduction to

# PROBABILITY & STATISTICS

fourteenth edition



MENDENHALL

BEAVER

BEAVER

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit [www.cengage.com/highered](http://www.cengage.com/highered) to search by ISBN#, author, title, or keyword for materials in your areas of interest.

# Introduction to Probability and Statistics

14<sup>th</sup>

EDITION

**William Mendenhall, III**

**Robert J. Beaver**

University of California, Riverside, Emeritus

**Barbara M. Beaver**

University of California, Riverside, Emeritus





**Introduction to Probability and  
Statistics, Fourteenth Edition  
Mendenhall/Beaver/Beaver**

Editor in Chief: Michelle Julet  
Publisher: Richard Stratton  
Senior Sponsoring Editor: Molly Taylor  
Assistant Editor: Shaylin Walsh  
Editorial Assistant: Alexander Gontar  
Associate Media Editor: Andrew Coppola  
Marketing Director: Mande Eckersley  
Senior Marketing Manager: Barb Bartoszek  
Marketing Coordinator: Michael Ledesma  
Marketing Communications Manager:  
Mary Anne Payumo  
Content Project Manager: Jill Quinn  
Art Director: Linda Helcher  
Senior Manufacturing Print Buyer: Diane  
Gibbons  
Rights Acquisition Specialist: Shalice  
Shah-Caldwell  
Production Service: MPS Limited, a  
Macmillan Company  
Cover Designer: Rokusek Design  
Cover Image: Vera Volkova/©  
Shutterstock  
Compositor: MPS Limited, a Macmillan  
Company

© 2013, 2009 Brooks/Cole, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at  
**Cengage Learning Customer & Sales Support, 1-800-354-9706**

For permission to use material from this text or product,  
submit all requests online at **www.cengage.com/permissions**.

Further permissions questions can be emailed to  
**permissionrequest@cengage.com**.

Library of Congress Control Number: 2011933688

Student Edition

ISBN-13: 978-1-133-10375-2

ISBN-10: 1-133-10375-8

**Brooks/Cole**

20 Channel Center Street  
Boston, MA 02210  
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil and Japan. Locate your local office at  
**international.cengage.com/region**

Cengage Learning products are represented in Canada by  
Nelson Education, Ltd.

For your course and learning solutions, visit  
**www.cengage.com**.

Purchase any of our products at your local college store  
or at our preferred online store **www.cengagebrain.com**.  
**Instructors:** Please visit **login.cengage.com** and log in to  
access instructor-specific resources

# Preface

---

Every time you pick up a newspaper or a magazine, watch TV, or surf the Internet, you encounter statistics. Every time you fill out a questionnaire, register at an online website, or pass your grocery rewards card through an electronic scanner, your personal information becomes part of a database containing your personal statistical information. You cannot avoid the fact that in this information age, data collection and analysis are an integral part of our day-to-day activities. In order to be an educated consumer and citizen, you need to understand how statistics are used and misused in our daily lives.

## THE SECRET TO OUR SUCCESS

The first college course in introductory statistics that we ever took used *Introduction to Probability and Statistics* by William Mendenhall. Since that time, this text—currently in the fourteenth edition—has helped several generations of students understand what statistics is all about and how it can be used as a tool in their particular area of application. The secret to the success of *Introduction to Probability and Statistics* is its ability to blend the old with the new. With each revision we try to build on the strong points of previous editions, while always looking for new ways to motivate, encourage, and interest students using new technological tools.

## HALLMARK FEATURES OF THE FOURTEENTH EDITION

The fourteenth edition retains the traditional outline for the coverage of descriptive and inferential statistics. This revision maintains the straightforward presentation of the thirteenth edition. In this spirit, we have continued to simplify and clarify the language and to make the language and style more readable and “user friendly”—without sacrificing the statistical integrity of the presentation. Great effort has been taken to explain not only how to apply statistical procedures, but also to explain

- how to meaningfully describe real sets of data
- what the results of statistical tests mean in terms of their practical applications
- how to evaluate the validity of the assumptions behind statistical tests
- what to do when statistical assumptions have been violated

## Exercises

In the tradition of all previous editions, the variety and number of real applications in the exercise sets is a major strength of this edition. We have revised the exercise sets to provide new and interesting real-world situations and real data sets, many of which are drawn from current periodicals and journals. The fourteenth edition contains over 1300 problems, many of which are new to this edition. A set of classic exercises compiled from previous editions is available on the website (<http://www.cengage.com/statistics/mendenhall>). Exercises are graduated in level of difficulty; some, involving only basic techniques, can be solved by almost all students, while others, involving practical applications and interpretation of results, will challenge students to use more sophisticated statistical reasoning and understanding.


## Organization and Coverage

We believe that Chapters 1 through 10—with the possible exception of Chapter 3—should be covered in the order presented. The remaining chapters can be covered in any order. The analysis of variance chapter precedes the regression chapter, so that the instructor can present the analysis of variance as part of a regression analysis. Thus, the most effective presentation would order these three chapters as well.


Chapters 1–3 present descriptive data analysis for both one and two variables, using both *MINITAB* and Microsoft Excel<sup>®</sup> graphics. Chapter 4 includes a full presentation of probability and probability distributions. Three optional sections—Counting Rules, the Total Law of Probability, and Bayes’ Rule—are placed into the general flow of text, and instructors will have the option of complete or partial coverage. The sections that present event relations, independence, conditional probability, and the Multiplication Rule have been rewritten in an attempt to clarify concepts that often are difficult for students to grasp. As in the thirteenth edition, the chapters on analysis of variance and linear regression include both calculational formulas and computer printouts in the basic text presentation. These chapters can be used with equal ease by instructors who wish to use the “hands-on” computational approach to linear regression and ANOVA and by those who choose to focus on the interpretation of computer-generated statistical printouts.

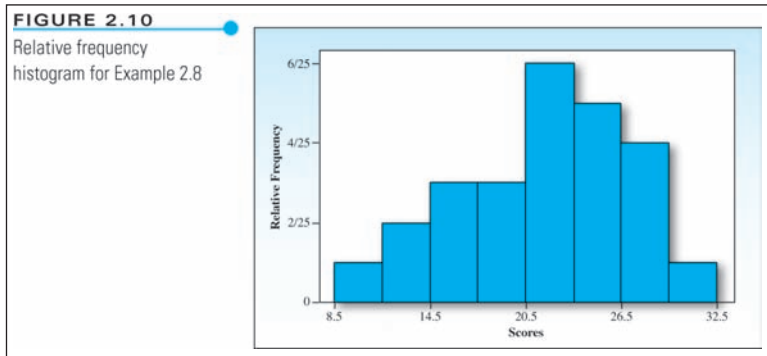
One important feature in the hypothesis testing chapters involves the emphasis on *p*-values and their use in judging statistical significance. With the advent of computer-generated *p*-values, these probabilities have become essential components in reporting the results of a statistical analysis. As such, the observed value of the test statistic and its *p*-value are presented together at the outset of our discussion of statistical hypothesis testing as equivalent tools for decision-making. Statistical significance is defined in terms of preassigned values of  $\alpha$ , and the *p*-value approach is presented as an alternative to the *critical value approach* for testing a statistical hypothesis. Examples are presented using both the *p*-value and *critical value* approaches to hypothesis testing. Discussion of the practical interpretation of statistical results, along with the difference between statistical significance and practical significance, is emphasized in the practical examples in the text.

## Special Features of the Fourteenth Edition

- **NEED TO KNOW. . .**: A special feature of this edition are highlighted sections called “NEED TO KNOW. . .” and identified by this icon.  **NEED TO KNOW...** These sections provide information consisting of definitions, procedures or step-by-step

hints on problem solving for specific questions such as “NEED TO KNOW... How to Construct a Relative Frequency Histogram?” or “NEED TO KNOW... How to Decide Which Test to Use?”

- Applets: Easy access to the Internet has made it possible for students to visualize statistical concepts using an interactive webtool called an applet. Applets written by Gary McClelland, author of *Seeing Statistics*<sup>TM</sup>, are found on the CourseMate  Website that accompanies the text. Following each applet, appropriate exercises are available that provide visual reinforcement of the concepts presented in the text. Applets allow the user to perform a statistical experiment, to interact with a statistical graph, to change its form, or to access an interactive “statistical table.”
- Graphical and numerical data description includes both traditional and EDA methods, using computer graphics generated by *MINITAB 16* for Windows and MS Excel.



**FIGURE 2.18** (b)

|                    | E                     | F      | G                    | H      |
|--------------------|-----------------------|--------|----------------------|--------|
|                    | <i>Front Leg Room</i> |        | <i>Rear Leg Room</i> |        |
| Mean               |                       | 41.333 | Mean                 | 28.944 |
| Standard Error     |                       | 0.344  | Standard Error       | 0.543  |
| Median             |                       | 41     | Median               | 29.5   |
| Mode               |                       | 41     | Mode                 | 30     |
| Standard Deviation |                       | 1.031  | Standard Deviation   | 1.629  |
| Sample Variance    |                       | 1.063  | Sample Variance      | 2.653  |
| Kurtosis           |                       | -0.745 | Kurtosis             | 1.735  |
| Skewness           |                       | 0.245  | Skewness             | -1.212 |
| Range              |                       | 3      | Range                | 5.5    |
| Minimum            |                       | 40     | Minimum              | 25.5   |
| Maximum            |                       | 43     | Maximum              | 31     |
| Sum                |                       | 372    | Sum                  | 260.5  |
| Count              |                       | 9      | Count                | 9      |

- All examples and exercises in the text contain printouts based on *MINITAB 16* and consistent with earlier versions of *MINITAB* or MS Excel. Printouts are provided for some exercises, while other exercises require the student to obtain solutions without using a computer.

**Data set** **1.47 Presidential Vetoes** Here is a list of the 44 presidents of the United States along with the number of regular vetoes used by each:<sup>5</sup>

|                |     |                 |     |
|----------------|-----|-----------------|-----|
| Washington     | 2   | B. Harrison     | 19  |
| J. Adams       | 0   | Cleveland       | 42  |
| Jefferson      | 0   | McKinley        | 6   |
| Madison        | 5   | T. Roosevelt    | 42  |
| Monroe         | 1   | Taft            | 30  |
| J. Q. Adams    | 0   | Wilson          | 33  |
| Jackson        | 5   | Harding         | 5   |
| Van Buren      | 0   | Coolidge        | 20  |
| W. H. Harrison | 0   | Hoover          | 21  |
| Tyler          | 6   | F. D. Roosevelt | 372 |
| Polk           | 2   | Truman          | 180 |
| Taylor         | 0   | Eisenhower      | 73  |
| Fillmore       | 0   | Kennedy         | 12  |
| Pierce         | 9   | L. Johnson      | 16  |
| Buchanan       | 4   | Nixon           | 26  |
| Lincoln        | 2   | Ford            | 48  |
| A. Johnson     | 21  | Carter          | 13  |
| Grant          | 45  | Reagan          | 39  |
| Hayes          | 12  | G. H. W. Bush   | 29  |
| Garfield       | 0   | Clinton         | 36  |
| Arthur         | 4   | G. W. Bush      | 11  |
| Cleveland      | 304 | Obama           | 1   |

Source: The World Almanac and Book of Facts 2011

Use an appropriate graph to describe the number of vetoes cast by the 44 presidents. Write a summary paragraph describing this set of data.

|        |       |        |        |                    |        |        |        |        |        |        |
|--------|-------|--------|--------|--------------------|--------|--------|--------|--------|--------|--------|
| (1950) | 121.3 | 122.3  | 121.3  | 122.0              | 123.0  | 121.4  | 123.2  | 122.1  | 125.0  | 122.1  |
| (1960) | 122.2 | 124.0  | 120.2  | 121.4              | 120.0  | 121.1  | 122.0  | 120.3  | 122.1  | 121.4  |
| (1970) | 123.2 | 123.1  | 121.4  | 119.2 <sup>†</sup> | 124.0  | 122.0  | 121.3  | 122.1  | 121.1  | 122.2  |
| (1980) | 122.0 | 122.0  | 122.2  | 122.1              | 122.2  | 120.1  | 122.4  | 123.2  | 122.2  | 125.0  |
| (1990) | 122.0 | 123.0  | 123.0  | 122.2              | 123.3  | 121.1  | 121.0  | 122.4  | 122.2  | 123.2  |
| (2000) | 121.0 | 119.97 | 121.13 | 121.19             | 124.06 | 122.75 | 121.36 | 122.17 | 121.86 | 122.66 |
| (2010) | 124.4 |        |        |                    |        |        |        |        |        |        |

<sup>†</sup>Record time set by Secretariat in 1973.  
Source: www.kentuckyderby.com

**a.** Do you think there will be a trend in the winning times over the years? Draw a line chart to verify your answer.

**b.** Describe the distribution of winning times using an appropriate graph. Comment on the shape of the distribution and look for any unusual observations.

---

**Data set** **1.48 Windy Cities** Are some cities more windy than others? Does Chicago deserve to be

**Data set** **1.50 Gulf Oil Spill Cleanup** On April 20, 2010, the United States experienced a major environmental disaster when a Deepwater Horizon drilling rig exploded in the Gulf of Mexico. The number of personnel and equipment used in the Gulf oil spill cleanup, beginning May 2, 2010 (Day 13) through June 9, 2010 (Day 51) is given in the following table.<sup>13</sup>

|                                   | Day 13 | Day 26 | Day 39 | Day 51 |
|-----------------------------------|--------|--------|--------|--------|
| Number of personnel (1000s)       | 3.0    | 17.5   | 20.0   | 24.0   |
| Federal Gulf fishing areas closed | 3%     | 8%     | 25%    | 32%    |
| Booms laid (miles)                | 46     | 315    | 644    | 909    |
| Dispersants used (1000 gallons)   | 156    | 500    | 870    | 1143   |

**TECHNOLOGY TODAY**

## The Role of Computers in the Fourteenth Edition—TECHNOLOGY TODAY

Computers are now a common tool for college students in all disciplines. Most students are accomplished users of word processors, spreadsheets, and databases, and they have no trouble navigating through software packages in the Windows environment. We believe, however, that advances in computer technology should not turn statistical analyses into a “black box.” Rather, we choose to use the computational shortcuts and interactive visual tools that modern technology provides to give us more time to emphasize statistical reasoning as well as the understanding and interpretation of statistical results.

In this edition, students will be able to use computers for both standard statistical analyses and as a tool for reinforcing and visualizing statistical concepts. Both MS Excel and *MINITAB 16* (consistent with earlier versions of *MINITAB*) are used exclusively as the computer packages for statistical analysis. However, we have chosen to isolate the instructions for generating computer output into individual sections called Technology Today at the end of each chapter. Each discussion uses numerical examples to guide the student through the MS Excel commands and option necessary for the procedures presented in that chapter, and then present the equivalent steps and commands needed to produce the same or similar results using *MINITAB*. We have included screen captures from both MS Excel and *MINITAB 16*, so that the student can actually work through these sections as “mini-labs.”

If you do not need “hands-on” knowledge of *MINITAB* or MS Excel, or if you are using another software package, you may choose to skip these sections and simply use the printouts as guides for the basic understanding of computer printouts.





**TECHNOLOGY TODAY**

**Numerical Descriptive Measures in Excel**

*MS Excel* provides most of the basic descriptive statistics presented in Chapter 2 using a single command on the **Data** tab. Other descriptive statistics can be calculated using the **Function** command on the **Formulas** tab.

**EXAMPLE**

2.15

The following data are the front and rear leg rooms (in inches) for nine different utility vehicles:<sup>14</sup>

| Make & Model              | Front Leg Room | Rear Leg Room |
|---------------------------|----------------|---------------|
| Acura MDX                 | 41.0           | 28.5          |
| Buick Enclave             | 41.5           | 30.0          |
| Chevy TrailBlazer         | 40.0           | 25.5          |
| Chevy Tahoe Hybrid V8 CVT | 41.0           | 27.5          |
| GMC Terrain 1LT 4-cyl     | 43.0           | 31.0          |

**Numerical Descriptive Measures in MINITAB**

*MINITAB* provides most of the basic descriptive statistics presented in Chapter 2 using a single command in the drop-down menus.

The following data are the front and rear leg rooms (in inches) for nine different sports utility vehicles:<sup>14</sup>

| Make and Model            | Front Leg Room | Rear Leg Room |
|---------------------------|----------------|---------------|
| Acura MDX                 | 41.0           | 28.5          |
| Buick Enclave             | 41.5           | 30.0          |
| Chevy TrailBlazer         | 40.0           | 25.5          |
| Chevy Tahoe Hybrid V8 CVT | 41.0           | 27.5          |
| GMC Terrain 1LT 4-cyl     | 43.0           | 31.0          |
| Honda CR-V                | 41.0           | 29.5          |
| Honda CR-V                | 40.5           | 28.5          |

Any student who has Internet access can use the applets found on the CourseMate Website to visualize a variety of statistical concepts (access instructions for the CourseMate Website are listed on the Printed Access Card that is an optional bundle with this text). In addition, some of the applets can be used instead of computer software to perform simple statistical analyses. Exercises written specifically for use with these applets also appear on the CourseMate Website. Students can use the applets at home or in a computer lab. They can use them as they read through the text material, once they have finished reading the entire chapter, or as a tool for exam review. Instructors can use the applets as a tool in a lab setting, or use them for visual demonstrations during lectures. We believe that these applets will be a powerful tool that will increase student enthusiasm for, and understanding of, statistical concepts and procedures.

**STUDY AIDS**

The many and varied exercises in the text provide the best learning tool for students embarking on a first course in statistics. The answers to all odd-numbered exercises are given in the back of the text, and a detailed solution appears in the *Student Solutions Manual*, which is available as a supplement for students. Each application exercise has

a title, making it easier for students and instructors to immediately identify both the context of the problem and the area of application.

Use Table 2 to find the following probabilities:

- a.  $P(x \leq 3)$                       b.  $P(x > 3)$   
 c.  $P(x = 3)$                       d.  $P(3 \leq x \leq 5)$

**5.38** Consider a Poisson random variable with  $\mu = 0.8$ .

Use Table 2 to find the following probabilities:

- a.  $P(x = 0)$                       b.  $P(x \leq 2)$   
 c.  $P(x > 2)$                       d.  $P(2 \leq x \leq 4)$

**5.39** Let  $x$  be a Poisson random variable with mean  $\mu = 2$ . Calculate these probabilities:

- a.  $P(x = 0)$                       b.  $P(x = 1)$   
 c.  $P(x > 1)$                       d.  $P(x = 5)$

## APPLICATIONS

**5.43 Airport Safety** The increased number of small commuter planes in major airports has heightened concern over air safety. An eastern airport has recorded a monthly average of five near misses on landings and takeoffs in the past 5 years.

- Find the probability that during a given month there are no near misses on landings and takeoffs at the airport.
- Find the probability that during a given month there are five near misses.
- Find the probability that there are at least five near-

Students should be encouraged to use the “NEED TO KNOW. . .” sections as they occur in the text. The placement of these sections is intended to answer questions as they would normally arise in discussions. In addition, there are numerous hints called “NEED A TIP?” that appear in the margins of the text. The tips are short and concise.

### NEED A TIP

Empirical Rule  $\leftrightarrow$   
 mound-shaped data  
 Tchebysheff  $\leftrightarrow$  any  
 shaped data

Is Tchebysheff’s Theorem applicable? Yes, because it can be used for any set of data. According to Tchebysheff’s Theorem,

- at least  $3/4$  of the measurements will fall between 10.6 and 32.6.
- at least  $8/9$  of the measurements will fall between 5.1 and 38.1.

Finally, sections called **Key Concepts and Formulas** appear in each chapter as a review in outline form of the material covered in that chapter.

## CHAPTER REVIEW

### Key Concepts and Formulas

#### I. Measures of the Center of a Data Distribution

1. Arithmetic mean (mean) or average

- Population:  $\mu$
- Sample of  $n$  measurements:  $\bar{x} = \frac{\sum x_i}{n}$

2. Median; **position** of the median =  $.5(n + 1)$

3. Mode

4. The median may be preferred to the mean if the data are highly skewed.

#### II. Measures of Variability

1. Range:  $R = \text{largest} - \text{smallest}$

2. Variance

a. Population of  $N$  measurements:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

b. Sample of  $n$  measurements:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1}$$

68%, 95%, and 99.7% of the measurements are within one, two, and three standard deviations of the mean, respectively.

#### IV. Measures of Relative Standing

- Sample  $z$ -score:  $z = \frac{x - \bar{x}}{s}$
- $p$ th percentile;  $p\%$  of the measurements are smaller, and  $(100 - p)\%$  are larger.
- Lower quartile,  $Q_1$ ; **position** of  $Q_1 = .25(n + 1)$
- Upper quartile,  $Q_3$ ; **position** of  $Q_3 = .75(n + 1)$
- Interquartile range:  $\text{IQR} = Q_3 - Q_1$


#### V. The Five-Number Summary and Box Plots

1. The **five-number summary**:

Min  $Q_1$  Median  $Q_3$  Max

One-fourth of the measurements in the data set lie between each of the four adjacent pairs of numbers.

2. Box plots are used for detecting outliers and

The CourseMate Website, a password-protected resource that can be accessed with a Printed Access Card (optional bundle item), provides students with an array of study resources, including the complete set of Java applets, the TI Calculator Tech Guide that includes instructions for performing many of the techniques in the text using the popular TI 83/84/89 graphing calculator, an interactive eBook, online Quizzes, flashcards, and more. The data sets (saved in a variety of formats) can be found on the book's website ([www.CengageBrain.com](http://www.CengageBrain.com)) as well as the CourseMate Website. 

## INSTRUCTOR RESOURCES

The **Instructor's Website** (<http://www.cengage.com/statistics/mendenhall>), available to adopters of the fourteenth edition, provides a variety of teaching aids, including

- All the material from the CourseMate website including exercises using the Large Data Sets, which is accompanied by three large data sets that can be used throughout the course. A file named “Fortune” contains the revenues (in millions) for the *Fortune* 500 largest U.S. industrial corporations in a recent year; a file named “Batting” contains the batting averages for the National and American baseball league batting champions from 1976 to 2010; and a file named “Blood Pressure” contains the age and diastolic and systolic blood pressures for 965 men and 945 women compiled by the National Institutes of Health.
- Classic exercises with data sets and solutions
- PowerPoint lecture slides
- Applets by Gary McClelland (the complete set of Java applets used for the MyApps exercises on the website)
- TI Calculator Tech Guide, which includes instructions for performing many of the techniques in the text using the TI-83/84/89 graphing calculators.

Also available for instructors:

### Aplia

Aplia is a web-based learning solution that increases student effort and engagement. It helps make statistics relevant and engaging to students by connecting real-world examples to course concepts. When combined with the textual material of *Introduction to Probability and Statistics* (IPS) 14,

- Students receive immediate, detailed explanations for every answer.
- Math and graphing tutorials help students overcome deficiencies in these crucial areas.
- Grades are automatically recorded in the instructor's Aplia gradebook.

### Solution Builder

This online instructor database offers complete worked-out solutions to all exercises in the text, allowing you to create customized, secure solutions printouts (in PDF format) matched exactly to the problems you assign in class. Sign up for access at [www.cengage.com/solutionbuilder](http://www.cengage.com/solutionbuilder).

### **PowerLecture™**

PowerLecture with ExamView® for *Introduction to Probability and Statistics* contains the Instructor's Solutions Manual, PowerPoint lectures, ExamView Computerized Testing, Classic Exercises, and TI-83/84/89 calculator Tech Guide which includes instructions for performing many of the techniques in the text using the TI-83/84/89 graphing calculators.

## **ACKNOWLEDGMENTS**

The authors are grateful to Molly Taylor and the editorial staff of Cengage Learning for their patience, assistance, and cooperation in the preparation of this edition. A special thanks to Gary McClelland for the Java applets used in the text.

Thanks are also due to fourteenth edition reviewers Ronald C. Degges, Bob C. Denton, Dr. Dorothy M. French, Jungwon Mun, Kazuhiko Shinki, Florence P. Shu and thirteenth edition reviewers Bob Denton, Timothy Husband, Rob LaBorde, Craig McBride, Marc Sylvester, Kanapathi Thiru, and Vitaly Voloshin. We wish to thank authors and organizations for allowing us to reprint selected material; acknowledgments are made wherever such material appears in the text.

*Robert J. Beaver*

*Barbara M. Beaver*

# Brief Contents

|           |                                                  |            |
|-----------|--------------------------------------------------|------------|
|           | <b>INTRODUCTION</b>                              | <b>1</b>   |
| <b>1</b>  | <b>DESCRIBING DATA WITH GRAPHS</b>               | <b>7</b>   |
| <b>2</b>  | <b>DESCRIBING DATA WITH NUMERICAL MEASURES</b>   | <b>50</b>  |
| <b>3</b>  | <b>DESCRIBING BIVARIATE DATA</b>                 | <b>94</b>  |
| <b>4</b>  | <b>PROBABILITY AND PROBABILITY DISTRIBUTIONS</b> | <b>123</b> |
| <b>5</b>  | <b>SEVERAL USEFUL DISCRETE DISTRIBUTIONS</b>     | <b>175</b> |
| <b>6</b>  | <b>THE NORMAL PROBABILITY DISTRIBUTION</b>       | <b>209</b> |
| <b>7</b>  | <b>SAMPLING DISTRIBUTIONS</b>                    | <b>242</b> |
| <b>8</b>  | <b>LARGE-SAMPLE ESTIMATION</b>                   | <b>281</b> |
| <b>9</b>  | <b>LARGE-SAMPLE TESTS OF HYPOTHESES</b>          | <b>324</b> |
| <b>10</b> | <b>INFERENCE FROM SMALL SAMPLES</b>              | <b>364</b> |
| <b>11</b> | <b>THE ANALYSIS OF VARIANCE</b>                  | <b>425</b> |
| <b>12</b> | <b>LINEAR REGRESSION AND CORRELATION</b>         | <b>482</b> |
| <b>13</b> | <b>MULTIPLE REGRESSION ANALYSIS</b>              | <b>530</b> |
| <b>14</b> | <b>ANALYSIS OF CATEGORICAL DATA</b>              | <b>574</b> |
| <b>15</b> | <b>NONPARAMETRIC STATISTICS</b>                  | <b>606</b> |
|           | <b>APPENDIX I</b>                                | <b>655</b> |
|           | <b>DATA SOURCES</b>                              | <b>688</b> |
|           | <b>ANSWERS TO SELECTED EXERCISES</b>             | <b>700</b> |
|           | <b>INDEX</b>                                     | <b>714</b> |

# Contents

## **Introduction: What is Statistics? 1**

- The Population and the Sample 3
- Descriptive and Inferential Statistics 4
- Achieving the Objective of Inferential Statistics: The Necessary Steps 4
- Keys for Successful Learning 5

1

## **DESCRIBING DATA WITH GRAPHS 7**

- 1.1 Variables and Data 8
- 1.2 Types of Variables 9
- 1.3 Graphs for Categorical Data 11
  - Exercises 14
- 1.4 Graphs for Quantitative Data 17
  - Pie Charts and Bar Charts 17
  - Line Charts 19
  - Dotplots 20
  - Stem and Leaf Plots 20
  - Interpreting Graphs with a Critical Eye 22
- 1.5 Relative Frequency Histograms 24
  - Exercises 28
  - Chapter Review 33**
  - Technology Today 33**
  - Supplementary Exercises 42**
  - CASE STUDY: How Is Your Blood Pressure? 49**

2

## **DESCRIBING DATA WITH NUMERICAL MEASURES 50**

- 2.1 Describing a Set of Data with Numerical Measures 51
- 2.2 Measures of Center 51
  - Exercises 55
- 2.3 Measures of Variability 57
  - Exercises 62

- 2.4 On the Practical Significance of the Standard Deviation 63
- 2.5 A Check on the Calculation of  $s$  67
  - Exercises 69
- 2.6 Measures of Relative Standing 72
- 2.7 The Five-Number Summary and the Box Plot 77
  - Exercises 80
  - Chapter Review 83**
  - Technology Today 84**
  - Supplementary Exercises 87**
  - CASE STUDY: The Boys of Summer 93**

### 3 DESCRIBING BIVARIATE DATA 94

- 3.1 Bivariate Data 95
- 3.2 Graphs for Categorical Variables 95
  - Exercises 98
- 3.3 Scatterplots for Two Quantitative Variables 99
- 3.4 Numerical Measures for Quantitative Bivariate Data 101
  - Exercises 107
  - Chapter Review 109**
  - Technology Today 109**
  - Supplementary Exercises 114**
  - CASE STUDY: Are Your Dishes *Really* Clean? 121**

### 4 PROBABILITY AND PROBABILITY DISTRIBUTIONS 123

- 4.1 The Role of Probability in Statistics 124
- 4.2 Events and the Sample Space 124
- 4.3 Calculating Probabilities Using Simple Events 127
  - Exercises 130
- 4.4 Useful Counting Rules (Optional) 133
  - Exercises 137
- 4.5 Event Relations and Probability Rules 139
  - Calculating Probabilities for Unions and Complements 141
- 4.6 Independence, Conditional Probability, and the Multiplication Rule 144
  - Exercises 149
- 4.7 Bayes' Rule (Optional) 152
  - Exercises 156

|     |                                                                 |            |
|-----|-----------------------------------------------------------------|------------|
| 4.8 | Discrete Random Variables and Their Probability Distributions   | 158        |
|     | Random Variables                                                | 158        |
|     | Probability Distributions                                       | 158        |
|     | The Mean and Standard Deviation for a Discrete Random Variable  | 160        |
|     | Exercises                                                       | 163        |
|     | <b>Chapter Review</b>                                           | <b>166</b> |
|     | <b>Technology Today</b>                                         | <b>167</b> |
|     | <b>Supplementary Exercises</b>                                  | <b>169</b> |
|     | <b>CASE STUDY: Probability and Decision Making in the Congo</b> | <b>174</b> |

5

## SEVERAL USEFUL DISCRETE DISTRIBUTIONS 175

|     |                                                      |            |
|-----|------------------------------------------------------|------------|
| 5.1 | Introduction                                         | 176        |
| 5.2 | The Binomial Probability Distribution                | 176        |
|     | Exercises                                            | 185        |
| 5.3 | The Poisson Probability Distribution                 | 188        |
|     | Exercises                                            | 193        |
| 5.4 | The Hypergeometric Probability Distribution          | 194        |
|     | Exercises                                            | 196        |
|     | <b>Chapter Review</b>                                | <b>197</b> |
|     | <b>Technology Today</b>                              | <b>198</b> |
|     | <b>Supplementary Exercises</b>                       | <b>202</b> |
|     | <b>CASE STUDY: A Mystery: Cancers Near a Reactor</b> | <b>208</b> |

6

## THE NORMAL PROBABILITY DISTRIBUTION 209

|     |                                                                              |            |
|-----|------------------------------------------------------------------------------|------------|
| 6.1 | Probability Distributions for Continuous Random Variables                    | 210        |
| 6.2 | The Normal Probability Distribution                                          | 213        |
| 6.3 | Tabulated Areas of the Normal Probability Distribution                       | 214        |
|     | The Standard Normal Random Variable                                          | 214        |
|     | Calculating Probabilities for a General Normal Random Variable               | 218        |
|     | Exercises                                                                    | 221        |
| 6.4 | The Normal Approximation to the Binomial Probability Distribution (Optional) | 224        |
|     | Exercises                                                                    | 229        |
|     | <b>Chapter Review</b>                                                        | <b>231</b> |
|     | <b>Technology Today</b>                                                      | <b>232</b> |
|     | <b>Supplementary Exercises</b>                                               | <b>236</b> |
|     | <b>CASE STUDY: "Are You Going to Curve the Grades?"</b>                      | <b>241</b> |



7

**SAMPLING DISTRIBUTIONS 242**

- 7.1 Introduction 243
- 7.2 Sampling Plans and Experimental Designs 243
  - Exercises 246
- 7.3 Statistics and Sampling Distributions 248
- 7.4 The Central Limit Theorem 251
- 7.5 The Sampling Distribution of the Sample Mean 254
  - Standard Error 255
  - Exercises 258
- 7.6 The Sampling Distribution of the Sample Proportion 260
  - Exercises 264
- 7.7 A Sampling Application: Statistical Process Control (Optional) 266
  - A Control Chart for the Process Mean: The  $\bar{x}$  Chart 267
  - A Control Chart for the Proportion Defective: The  $p$  Chart 269
  - Exercises 271
  - Chapter Review 272**
  - Technology Today 273**
  - Supplementary Exercises 276**
  - CASE STUDY: Sampling the Roulette at Monte Carlo 279**

8

**LARGE-SAMPLE ESTIMATION 281**

- 8.1 Where We've Been 282
- 8.2 Where We're Going—Statistical Inference 282
- 8.3 Types of Estimators 283
- 8.4 Point Estimation 284
  - Exercises 289
- 8.5 Interval Estimation 291
  - Constructing a Confidence Interval 292
  - Large-Sample Confidence Interval for a Population Mean  $\mu$  294
  - Interpreting the Confidence Interval 295
  - Large-Sample Confidence Interval for a Population Proportion  $p$  297
  - Exercises 299
- 8.6 Estimating the Difference between Two Population Means 301
  - Exercises 304
- 8.7 Estimating the Difference between Two Binomial Proportions 307
  - Exercises 309
- 8.8 One-Sided Confidence Bounds 311

- 8.9 Choosing the Sample Size 312  
 Exercises 316  
**Chapter Review 318**  
**Supplementary Exercises 318**  
**CASE STUDY: How Reliable Is That Poll?**  
***CBS News: How and Where America Eats* 322**

## 9 LARGE-SAMPLE TESTS OF HYPOTHESES 324

- 9.1 Testing Hypotheses about Population Parameters 325  
 9.2 A Statistical Test of Hypothesis 325  
 9.3 A Large-Sample Test about a Population Mean 328  
 The Essentials of the Test 329  
 Calculating the  $p$ -Value 332  
 Two Types of Errors 335  
 The Power of a Statistical Test 336  
 Exercises 339  
 9.4 A Large-Sample Test of Hypothesis for the Difference between Two Population Means 341  
 Hypothesis Testing and Confidence Intervals 343  
 Exercises 344  
 9.5 A Large-Sample Test of Hypothesis for a Binomial Proportion 347  
 Statistical Significance and Practical Importance 349  
 Exercises 350  
 9.6 A Large-Sample Test of Hypothesis for the Difference between Two Binomial Proportions 351  
 Exercises 354  
 9.7 Some Comments on Testing Hypotheses 356  
**Chapter Review 357**  
**Supplementary Exercises 358**  
**CASE STUDY: An Aspirin a Day . . . ? 362**

## 10 INFERENCE FROM SMALL SAMPLES 364

- 10.1 Introduction 365  
 10.2 Student's  $t$  Distribution 365  
 Assumptions behind Student's  $t$  Distribution 368  
 10.3 Small-Sample Inferences Concerning a Population Mean 369  
 Exercises 373  
 10.4 Small-Sample Inferences for the Difference between Two Population Means: Independent Random Samples 376  
 Exercises 382

10.5 Small-Sample Inferences for the Difference between  
Two Means: A Paired-Difference Test 386

Exercises 391

10.6 Inferences Concerning a Population Variance 394

Exercises 400

10.7 Comparing Two Population Variances 401

Exercises 407

10.8 Revisiting the Small-Sample Assumptions 409

**Chapter Review 410**

**Technology Today 410**

**Supplementary Exercises 416**

**CASE STUDY: School Accountability Study—  
How Is Your School Doing? 424**

**11 THE ANALYSIS OF VARIANCE 425**

11.1 The Design of an Experiment 426

11.2 What Is an Analysis of Variance? 427

11.3 The Assumptions for an Analysis of Variance 427

11.4 The Completely Randomized Design: A One-Way Classification 428

11.5 The Analysis of Variance for a Completely Randomized Design 429

Partitioning the Total Variation in an Experiment 429

Testing the Equality of the Treatment Means 432

Estimating Differences in the Treatment Means 434

Exercises 437

11.6 Ranking Population Means 440

Exercises 443

11.7 The Randomized Block Design: A Two-Way Classification 444

11.8 The Analysis of Variance for a Randomized Block Design 445

Partitioning the Total Variation in the Experiment 445

Testing the Equality of the Treatment and Block Means 448

Identifying Differences in the Treatment and Block Means 450

Some Cautionary Comments on Blocking 451

Exercises 452

11.9 The  $a \times b$  Factorial Experiment: A Two-Way Classification 456

11.10 The Analysis of Variance for an  $a \times b$  Factorial Experiment 458

Exercises 462

11.11 Revisiting the Analysis of Variance Assumptions 466

Residual Plots 467

11.12 A Brief Summary 469

**Chapter Review 469**

**Technology Today 470**

**Supplementary Exercises 475**

**CASE STUDY: How to Save Money on Groceries! 481**

12

## **LINEAR REGRESSION AND CORRELATION 482**

12.1 Introduction 483

12.2 A Simple Linear Probabilistic Model 483

12.3 The Method of Least Squares 486

12.4 An Analysis of Variance for Linear Regression 488

Exercises 491

12.5 Testing the Usefulness of the Linear Regression Model 494

Inferences Concerning  $\beta$ , the Slope of the Line of Means 495

The Analysis of Variance  $F$ -Test 498

Measuring the Strength of the Relationship:

The Coefficient of Determination 498

Interpreting the Results of a Significant Regression 499

Exercises 500

12.6 Diagnostic Tools for Checking the Regression Assumptions 503

Dependent Error Terms 503

Residual Plots 503

Exercises 504

12.7 Estimation and Prediction Using the Fitted Line 507

Exercises 511

12.8 Correlation Analysis 513

Exercises 517

**Chapter Review 519**

**Technology Today 520**

**Supplementary Exercises 523**

**CASE STUDY: Is Your Car “Made in the U.S.A.”? 528**

13

## **MULTIPLE REGRESSION ANALYSIS 530**

13.1 Introduction 531

13.2 The Multiple Regression Model 531

13.3 A Multiple Regression Analysis 532

The Method of Least Squares 533

The Analysis of Variance for Multiple Regression 534

Testing the Usefulness of the Regression Model 535

Interpreting the Results of a Significant Regression 536

|                                                                                          |            |
|------------------------------------------------------------------------------------------|------------|
| Checking the Regression Assumptions                                                      | 538        |
| Using the Regression Model for Estimation and Prediction                                 | 538        |
| <b>13.4 A Polynomial Regression Model</b>                                                | <b>539</b> |
| Exercises                                                                                | 542        |
| <b>13.5 Using Quantitative and Qualitative Predictor Variables in a Regression Model</b> | <b>546</b> |
| Exercises                                                                                | 552        |
| <b>13.6 Testing Sets of Regression Coefficients</b>                                      | <b>555</b> |
| <b>13.7 Interpreting Residual Plots</b>                                                  | <b>558</b> |
| <b>13.8 Stepwise Regression Analysis</b>                                                 | <b>559</b> |
| <b>13.9 Misinterpreting a Regression Analysis</b>                                        | <b>560</b> |
| Causality                                                                                | 560        |
| Multicollinearity                                                                        | 560        |
| <b>13.10 Steps to Follow When Building a Multiple Regression Model</b>                   | <b>562</b> |
| <b>Chapter Review</b>                                                                    | <b>562</b> |
| <b>Technology Today</b>                                                                  | <b>563</b> |
| <b>Supplementary Exercises</b>                                                           | <b>565</b> |
| <b>CASE STUDY: “Made in the U.S.A.”—Another Look</b>                                     | <b>572</b> |

14

**ANALYSIS OF CATEGORICAL DATA 574**

|                                                                                                          |            |
|----------------------------------------------------------------------------------------------------------|------------|
| 14.1 A Description of the Experiment                                                                     | 575        |
| 14.2 Pearson’s Chi-Square Statistic                                                                      | 576        |
| 14.3 Testing Specified Cell Probabilities: The Goodness-of-Fit Test                                      | 577        |
| Exercises                                                                                                | 579        |
| 14.4 Contingency Tables: A Two-Way Classification                                                        | 581        |
| The Chi-Square Test of Independence                                                                      | 582        |
| Exercises                                                                                                | 586        |
| 14.5 Comparing Several Multinomial Populations: A Two-Way Classification with Fixed Row or Column Totals | 588        |
| Exercises                                                                                                | 591        |
| 14.6 The Equivalence of Statistical Tests                                                                | 592        |
| 14.7 Other Applications of the Chi-Square Test                                                           | 593        |
| <b>Chapter Review</b>                                                                                    | <b>594</b> |
| <b>Technology Today</b>                                                                                  | <b>595</b> |
| <b>Supplementary Exercises</b>                                                                           | <b>598</b> |
| <b>CASE STUDY: Who is the Primary Breadwinner in Your Family?</b>                                        | <b>604</b> |

**NONPARAMETRIC STATISTICS 606**

- 15.1 Introduction 607
- 15.2 The Wilcoxon Rank Sum Test: Independent Random Samples 607
  - Normal Approximation for the Wilcoxon Rank Sum Test 611
  - Exercises 614
- 15.3 The Sign Test for a Paired Experiment 616
  - Normal Approximation for the Sign Test 617
  - Exercises 619
- 15.4 A Comparison of Statistical Tests 620
- 15.5 The Wilcoxon Signed-Rank Test for a Paired Experiment 621
  - Normal Approximation for the Wilcoxon Signed-Rank Test 624
  - Exercises 625
- 15.6 The Kruskal–Wallis  $H$ -Test for Completely Randomized Designs 627
  - Exercises 631
- 15.7 The Friedman  $F_r$ -Test for Randomized Block Designs 633
  - Exercises 636
- 15.8 Rank Correlation Coefficient 637
  - Exercises 641
- 15.9 Summary 643
  - Chapter Review 644**
  - Technology Today 645**
  - Supplementary Exercises 648**
  - CASE STUDY: How’s Your Cholesterol Level? 653**

**APPENDIX I 655**

- Table 1 Cumulative Binomial Probabilities 656
- Table 2 Cumulative Poisson Probabilities 662
- Table 3 Areas under the Normal Curve 664
- Table 4 Critical Values of  $t$  667
- Table 5 Critical Values of Chi-Square 668
- Table 6 Percentage Points of the  $F$  Distribution 670
- Table 7 Critical Values of  $T$  for the Wilcoxon Rank Sum Test,  $n_1 \leq n_2$  678
- Table 8 Critical Values of  $T$  for the Wilcoxon Signed-Rank Test,  $n = 5(1)50$  680
- Table 9 Critical Values of Spearman’s Rank Correlation Coefficient for a One-Tailed Test 681

Table 10 Random Numbers 682

Table 11 Percentage Points of the Studentized Range,  $q_{.05}(k, df)$  684

**DATA SOURCES 688**

**ANSWERS TO SELECTED EXERCISES 700**

**INDEX 714**

*This page intentionally left blank*



# Introduction

## What is Statistics?

What is statistics? Have you ever met a statistician? Do you know what a statistician does? Perhaps you are thinking of the person who sits in the broadcast booth at the Rose Bowl, recording the number of pass completions, yards rushing, or interceptions thrown on New Year's Day. Or perhaps the mere mention of the word *statistics* sends a shiver of fear through you. You may think you know nothing about statistics; however, it is almost inevitable that you encounter statistics in one form or another every time you pick up a daily newspaper. Here are some examples concerning the California 2010 elections:

- **Rowdy crowd jeers Whitman.** GOP candidate criticizes unions; earlier stop draws friendlier audience.  
**GLENDALE**— . . . Whitman, a billionaire, has spent \$142 million from her personal fortune in the race so far. A Field Poll released Thursday showed her trailing Jerry Brown 49 percent to 39 percent among likely voters.<sup>1</sup>
- **Fiorina calls herself similar to Feinstein, who supports Boxer.**  
**MENLO PARK**—Republican Carly Fiorina said Friday she would be a like-minded colleague of Democratic Sen. Dianne Feinstein if she unseats Barbara Boxer next week, drawing sharp responses from both Democratic senators. . . . Fiorina, the former CEO of Hewlett-Packard Co., disputed a Field Poll released Friday showing Boxer leading her among likely voters, 49 percent to 41 percent.<sup>2</sup>
- **Race for attorney general tight. Field Poll:** Nearly a quarter of those surveyed are undecided. Newsom holds a slim lead over Maldonado for lieutenant governor.



© Mark Karrass/CORBIS

**SACRAMENTO**—Tuesday’s election for attorney general is a tossup, with Democrat Kamala Harris and Republican Steve Cooley virtually tied as Harris gains ground in voter-rich Los Angeles County and among women according to the latest Field Poll.

... Today’s poll shows Cooley with 39 percent and Harris with 38 percent among likely voters. Almost a quarter of likely voters remain undecided.

... Newsom, the mayor of San Francisco, leads Maldonado, who was appointed lieutenant governor this year, 42 percent to 37 percent. A fifth of voters are undecided.

Today’s poll was conducted for *The Press-Enterprise* and other California media subscribers. It was conducted October 14 through October 26 and included 1092 voters. It has a margin of error of plus or minus 3.2 percent.<sup>3</sup>

—*The Press-Enterprise*, Riverside, CA

Articles similar to these are commonplace in our newspapers and magazines, and in the period just prior to a presidential or congressional election, a new poll is reported almost every day. The language of these articles are very familiar to us; however, they leave the inquisitive reader with some unanswered questions. How were the people in the poll selected? Will these people give the same response tomorrow? Will they give the same response on election day? Will they even vote? Are these people representative of all those who will vote on election day? It is the job of a statistician to ask these questions and to find answers for them in the language of the poll.

#### Most Believe “Cover-Up” of JFK Assassination Facts

A majority of the public believes the assassination of President John F. Kennedy was part of a larger conspiracy, not the act of one individual. In addition, most Americans think there was a cover-up of facts about the 1963 shooting. Almost 50 years after JFK’s assassination, a FOX news poll shows many Americans disagree with the government’s conclusions about the killing. The **Warren Commission** found that **Lee Harvey Oswald** acted alone when he shot Kennedy, but 66 percent of the public today think the assassination was “part of a larger conspiracy” while only 25 percent think it was the “act of one individual.”

“For older Americans, the Kennedy assassination was a traumatic experience that began a loss of confidence in government,” commented Opinion Dynamics President John Gorman. “Younger people have grown up with movies and documentaries that have pretty much pushed the ‘conspiracy’ line. Therefore, it isn’t surprising there is a fairly solid national consensus that we still don’t know the truth.”

(The poll asked): “Do you think that we know all the facts about the assassination of President John F. Kennedy or do you think there was a cover-up?”

|                     | We Know All the Facts (%) | There Was a Cover-Up | (Not Sure) |
|---------------------|---------------------------|----------------------|------------|
| <b>All</b>          | 14                        | 74                   | 12         |
| <b>Democrats</b>    | 11                        | 81                   | 8          |
| <b>Republicans</b>  | 18                        | 69                   | 13         |
| <b>Independents</b> | 12                        | 71                   | 17         |

—www.foxnews.com<sup>4</sup>

When you see an article like this one in a magazine, do you simply read the title and the first paragraph, or do you read further and try to understand the meaning of the numbers? How did the authors get these numbers? Did they really interview every American with each political affiliation? It is the job of the statistician to interpret the language of this study.

#### Hot News: 98.6 Not Normal

After believing for more than a century that 98.6 was the normal body temperature for humans, researchers now say normal is not normal anymore.

For some people at some hours of the day, 99.9 degrees could be fine. And readings as low as 96 turn out to be highly human.

The 98.6 standard was derived by a German doctor in 1868. Some physicians have always been suspicious of the good doctor's research. His claim: 1 million readings—in an epoch without computers.

So Mackowiak & Co. took temperature readings from 148 healthy people over a three-day period and found that the mean temperature was 98.2 degrees. Only 8 percent of the readings were 98.6.

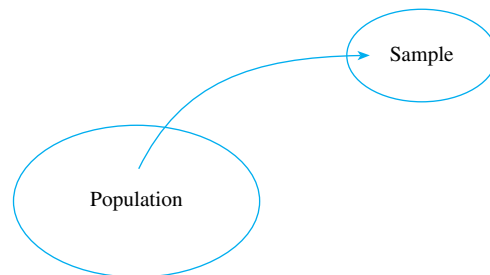
—*The Press-Enterprise*<sup>5</sup>

What questions come to your mind when you read this article? How did the researcher select the 148 people, and how can we be sure that the results based on these 148 people are accurate when applied to the general population? How did the researcher arrive at the normal “high” and “low” temperatures given in the article? How did the German doctor record 1 million temperatures in 1868? Again, we encounter a statistical problem with an application to everyday life.

Statistics is a branch of mathematics that has applications in almost every facet of our daily life. It is a new and unfamiliar language for most people, however, and, like any new language, statistics can seem overwhelming at first glance. But once the language of statistics is learned and understood, it provides a powerful tool for data analysis in many different fields of application.

## THE POPULATION AND THE SAMPLE

In the language of statistics, one of the most basic concepts is **sampling**. In most statistical problems, a specified number of measurements or data—a **sample**—is drawn from a much larger body of measurements, called the **population**.



For the body-temperature experiment, the sample is the set of body-temperature measurements for the 148 healthy people chosen by the experimenter. We hope that the sample is representative of a much larger body of measurements—the population—the body temperatures of all healthy people in the world!

Which is of primary interest, the sample or the population? In most cases, we are interested primarily in the population, but the population may be difficult or impossible to enumerate. Imagine trying to record the body temperature of every healthy person on earth or the presidential preference of every registered voter in the United States! Instead, **we try to describe or predict the behavior of the population on the basis of information obtained from a representative sample from that population.**

The words *sample* and *population* have two meanings for most people. For example, you read in the newspapers that a Gallup poll conducted in the United States was based on a sample of 1823 people. Presumably, each person interviewed is asked a particular question, and that person's response represents a single measurement in the sample. Is the sample the set of 1823 people, or is it the 1823 responses that they give?

In statistics, we distinguish between the set of objects on which the measurements are taken and the measurements themselves. To experimenters, the objects on which measurements are taken are called **experimental units**. The sample survey statistician calls them **elements of the sample**.

## DESCRIPTIVE AND INFERENCEAL STATISTICS

When first presented with a set of measurements—whether a sample or a population—you need to find a way to organize and summarize it. The branch of statistics that presents techniques for describing sets of measurements is called **descriptive statistics**. You have seen descriptive statistics in many forms: bar charts, pie charts, and line charts presented by a political candidate; numerical tables in the newspaper; or the average rainfall amounts reported by the local television weather forecaster. Computer-generated graphics and numerical summaries are commonplace in our everyday communication.

**Definition** **Descriptive statistics** consists of procedures used to summarize and describe the important characteristics of a set of measurements.

If the set of measurements is the entire population, you need only to draw conclusions based on the descriptive statistics. However, it might be too expensive or too time consuming to enumerate the entire population. Perhaps enumerating the population would destroy it, as in the case of “time to failure” testing. For these or other reasons, you may have only a sample from the population. By looking at the sample, you want to answer questions about the population as a whole. The branch of statistics that deals with this problem is called **inferential statistics**.

**Definition** **Inferential statistics** consists of procedures used to make inferences about population characteristics from information contained in a sample drawn from this population.

The **objective of inferential statistics** is to make inferences (that is, draw conclusions, make predictions, make decisions) about the characteristics of a population from information contained in a sample.

## ACHIEVING THE OBJECTIVE OF INFERENCEAL STATISTICS: THE NECESSARY STEPS

How can you make inferences about a population using information contained in a sample? The task becomes simpler if you organize the problem into a series of logical steps.

1. **Specify the questions to be answered and identify the population of interest.** In the California election poll, the objective is to determine who will get the most votes on election day. Hence, the population of interest is the set of all votes in the California election. When you select a sample, it is important that

the sample be representative of *this* population, not the population of voter preferences on October 30 or on some other day prior to the election.

2. **Decide how to select the sample.** This is called the *design of the experiment* or the *sampling procedure*. Is the sample representative of the population of interest? For example, if a sample of registered voters is selected from the city of San Francisco, will this sample be representative of all voters in California? Will it be the same as a sample of “likely voters”—those who are likely to actually vote in the election? Is the sample large enough to answer the questions posed in step 1 without wasting time and money on additional information? A good sampling design will answer the questions posed with minimal cost to the experimenter.
3. **Select the sample and analyze the sample information.** No matter how much information the sample contains, you must use an appropriate method of analysis to extract it. Many of these methods, which depend on the sampling procedure in step 2, are explained in the text.
4. **Use the information from step 3 to make an inference about the population.** Many different procedures can be used to make this inference, and some are better than others. For example, 10 different methods might be available to estimate human response to an experimental drug, but one procedure might be more accurate than others. You should use the best inference-making procedure available (many of these are explained in the text).
5. **Determine the reliability of the inference.** Since you are using only a fraction of the population in drawing the conclusions described in step 4, you might be wrong! How can this be? If an agency conducts a statistical survey for you and estimates that your company’s product will gain 34% of the market this year, how much confidence can you place in this estimate? Is this estimate accurate to within 1, 5, or 20 percentage points? Is it reliable enough to be used in setting production goals? Every statistical inference should include a measure of reliability that tells you how much confidence you have in the inference.

Now that you have learned a few basic terms and concepts, we again pose the question asked at the beginning of this discussion: Do you know what a statistician does? The statistician’s job is to implement all of the preceding steps.



## KEYS FOR SUCCESSFUL LEARNING

As you begin to study statistics, you will find that there are many new terms and concepts to be mastered. Since statistics is an applied branch of mathematics, many of these basic concepts are mathematical—developed and based on results from calculus or higher mathematics. However, you do not have to be able to derive results in order to apply them in a logical way. In this text, we use numerical examples and common-sense arguments to explain statistical concepts, rather than more complicated mathematical arguments.

In recent years, computers have become readily available to many students and provide them with an invaluable tool. In the study of statistics, even the beginning student can use packaged programs to perform statistical analyses with a high degree of speed and accuracy. Some of the more common statistical packages available at computer facilities are *MINITAB*<sup>TM</sup>, *SAS* (Statistical Analysis System), and *SPSS*