> STATISTICS

A Tool for Social Research

TENTH EDITION

JOSEPH F. HEALEY

STATISTICS

A Tool for Social Research

Tenth Edition

Joseph F. Healey

Christopher Newport University

CENGAGE Learning®

Australia • Brazil • Japan • Mexico • Singapore • United Kingdom • United States

CENGAGE Learning

Statistics: A Tool for Social Research, Tenth Edition
Joseph F. Healey

Product Manager: Seth Dobrin

Content Developer: Liana Sarkisian

Content Coordinator: Rachel McDonald

Product Assistant: Coco Bator

Media Developer: John Chell

Brand Manager: Kara Kindstrom

Content Project Manager: Cheri Palmer

Art Director: Caryl Gorska

Manufacturing Planner: Judy Inouye

Rights Acquisitions Specialist: Thomas McDonough

Production Service: Jill Traut, MPS Limited

Text Researcher: PreMediaGlobal

Copy Editor: Susan Reiland

Illustrator and Compositor: MPS Limited

Text Designer: Diane Beasley

Cover Designer: Lee Friedman

Cover Image: © Jurgen Ziewe/ Ikon Image/Corbis

Library of Congress Control Number: 2013952193

ISBN-13: 978-1-285-45885-4

ISBN-10: 1-285-45885-0

Cengage Learning
200 First Stamford Place, 4th Floor
Stamford, CT 06902
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 **www.cengage.com/global**.

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

To learn more about Cengage 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**.

Printed in the United States of America

1 2 3 4 5 6 7 18 17 16 15 14

Brief Contents

Detailed Contents

Preface

Statistics are part of the everyday language of sociology and other social sciences (including political science, social work, public administration, criminal justice, urban studies, and gerontology). These disciplines are research-based and routinely use statistics to express knowledge and to discuss theory and research. To join the conversations being conducted in these disciplines, you must be literate in the vocabulary of research, data analysis, and scientific thinking. Knowledge of statistics will enable you to understand the professional research literature, conduct quantitative research yourself, contribute to the growing body of social science knowledge, and reach your full potential as a social scientist.

Although essential, learning (and teaching) statistics can be a challenge. Students in social science statistics courses typically have a wide range of mathematical backgrounds and an equally diverse set of career goals. They are often puzzled about the relevance of statistics for them, and, not infrequently, there is some math anxiety to deal with. This text introduces statistical analysis for the social sciences while addressing these realities.

The text makes minimal assumptions about mathematical background (the ability to read a simple formula is sufficient preparation for virtually all of the material in the text), and a variety of special features help students analyze data successfully. The text has been written especially for sociology and social work programs but is sufficiently flexible to be used in any program with a social science base.

The text is written at an intermediate level and its goal is to show the relevance and value of statistics for the social sciences. I emphasize interpretation and understanding statistics in the context of social science research, but I have not sacrificed comprehensive coverage or statistical correctness. Mathematical explanations are kept at an elementary level, as is appropriate in a first exposure to social statistics. For example, I do not treat formal probability theory per se in the text.[1] Rather, the background necessary for an understanding of inferential statistics is introduced, informally and intuitively, in Chapters 5 and 6 while considering the concepts of the normal curve and the sampling distribution.

The text does not claim that statistics are "fun" or that the material can be mastered without considerable effort. At the same time, students are not overwhelmed with abstract proofs, formula derivations, or mathematical theory, which can needlessly frustrate the learning experience at this level.

[1]A presentation of probability is available at the Web site for this text for those who are interested.

Goal of the Text

The goal of this text is to develop basic statistical literacy. The statistically literate person understands and appreciates the role of statistics in the research process, is competent to perform basic calculations, and can read and appreciate the professional research literature in his or her field as well as any research reports he or she may encounter in everyday life. This goal has not changed since the first edition of this text. However, in recognition of the fact that "mere computation" has become less of a challenge in this high-tech age, this edition continues to increase the stress on interpretation and computer applications while deemphasizing computation. This will be apparent in several ways. For example, the feature called "Interpreting Statistics" has been updated. These noncomputational sections are included in about half of the chapters and present detailed examples of "what to say after the statistics have been calculated." They use real data and real research situations to illustrate the process of developing understanding, and they exemplify how statistics can be used to answer important questions. The issues addressed in these sections include recent changes in the American family, the gender gap in income, and the correlates of street crime.

Also, in recognition of the fact that modern technology has rendered hand calculation increasingly obsolete, the end-of-chapter problems feature smaller, easier-to-handle data sets, although some more challenging problems are also included. A new section, called "Using SPSS," has been added to most chapters to demonstrate how to use a computerized statistical package to produce statistics. The end-of-chapter problems now include SPSS-based exercises, and research projects using SPSS are included at the end of almost all chapters. To accommodate the increased use of SPSS, several new data sets have been added to the text and the General Social Survey data set has been updated to 2012.

A further goal of this text is to provide examples of everyday applications of statistics. Boxed features entitled "Statistics in Everyday Life" appear in each chapter and highlight the relevance of statistics in the real world.

This edition continues to focus on the development of basic statistical literacy, the three aspects of which provide a framework for discussing the additional features of this text.

1. **An Appreciation of Statistics.** A statistically literate person understands the relevance of statistics for social research, can analyze and interpret the meaning of a statistical test, and can select an appropriate statistic for a given purpose and a given set of data. This textbook develops these skills, within the constraints imposed by the introductory nature of the course, in several ways.
    - *The relevance of statistics.* Chapter 1 includes a discussion of the role of statistics in social research and stresses their usefulness as ways of analyzing and manipulating data and answering research questions. Throughout the text, each example problem is framed in the context of a research situation. A question is posed and then, with the aid of a statistic, answered. This central theme of usefulness is further reinforced by a series of "Applying Statistics" boxes, each of which illustrates some specific way statistics can be used to answer questions, and by the "Using Statistics" feature that opens every chapter.

End-of-chapter problems are labeled by the social science discipline from which they are drawn: SOC for sociology, SW for social work, PS for political science, CJ for criminal justice, PA for public administration, and GER for gerontology. Identifying problems with specific disciplines allows students to more easily see the relevance of statistics to their own academic interests. (Not incidentally, they will also see that the disciplines have a large subject matter in common.)

- *Interpreting statistics.* For most students, interpretation—saying what statistics mean—is a big challenge. The ability to interpret statistics can be developed only by exposure and experience. To provide exposure, I have been careful, in the example problems, to express the meaning of the statistic in terms of the original research question. To provide experience, the end-of-chapter problems call for an interpretation of the statistic calculated. To provide examples, many of the answers to odd-numbered computational problems in the back of the text are expressed in words as well as numbers. The "Interpreting Statistics" sections provide additional, detailed examples of how to express the meaning of statistics.
- *Using statistics: Ideas for research projects.* Appendix E offers ideas for independent data-analysis projects for students. The projects require students to use SPSS to analyze a data set. They can be assigned at intervals throughout the semester or at the end of the course. Each project provides an opportunity for students to practice and apply their statistical skills and, above all, to exercise their ability to understand and interpret the statistics they produce.

2. **Computational Competence.** Students should emerge from their first course in statistics with the ability to perform elementary forms of data analysis. While computers have made computation less of an issue today, computation is inseparable from statistics, so I have included a number of features to help students cope with these mathematical challenges.
    - *"One Step at a Time"* boxes for each statistic break down computation into individual steps for maximum clarity and ease.
    - *Extensive problem sets* are provided at the end of each chapter. For the most part, these problems use fictitious data and are designed for ease of computation.
    - *Cumulative exercises* are included after each of Parts I–IV to provide practice in choosing, computing, and analyzing statistics. These exercises present only data sets and research questions. Students must choose appropriate statistics as part of the exercise.
    - *Solutions* to odd-numbered computational problems are provided so that students may check their answers.
    - *SPSS* gives students access to the computational power of the computer. This is explained in more detail later.

3. **The Ability to Read the Professional Social Science Literature.** The statistically literate person can comprehend and critically appreciate research reports written by others. The development of this skill is a particular problem

at the introductory level because (1) the vocabulary of professional researchers is so much more concise than the language of the textbook, and (2) the statistics featured in the literature are more advanced than those covered at the introductory level. To help bridge this gap, I have included, beginning in Chapter 1, a series of boxes labeled "Reading Statistics." In each such box, I briefly describe the reporting style typically used for the statistic in question and try to alert students about what to expect when they approach the professional literature. These inserts have been updated in this edition and most include excerpts from the research literature that illustrate how statistics are actually applied and interpreted by social scientists.

Additional Features

A number of other features make the text more meaningful for students and more useful for instructors.

- *Readability and clarity.* The writing style is informal and accessible to students without ignoring the traditional vocabulary of statistics. Problems and examples have been written to maximize student interest and to focus on issues of concern and significance. For the more difficult material (such as hypothesis testing), students are first walked through an example problem before being confronted by formal terminology and concepts. Each chapter ends with a summary of major points and formulas and a glossary of important concepts. A list of frequently used formulas inside the front cover and a glossary of symbols inside the back cover can be used for quick reference.
- *Organization and coverage.* The text is divided into four parts, with most of the coverage devoted to univariate descriptive statistics, inferential statistics, and bivariate measures of association. The distinction between description and inference is introduced in the first chapter and maintained throughout the text. In selecting statistics for inclusion, I have tried to strike a balance between the essential concepts with which students must be familiar and the amount of material students can reasonably be expected to learn in their first (and perhaps only) statistics course, all the while bearing in mind that different instructors will naturally wish to stress different aspects of the subject. Thus, the text covers a full gamut of the usual statistics, with each chapter broken into subsections so that instructors may choose the particular statistics they wish to include.
- *Learning objectives.* Learning objectives are stated at the beginning of each chapter. These are intended to serve as "study guides" and to help students identify and focus on the most important material.
- *Review of mathematical skills.* A comprehensive review of all of the mathematical skills that will be used in this text is included as a Prologue. Students who are inexperienced or out of practice with mathematics can study this review early in the course and/or refer to it as needed. A self-test is included so that students can check their level of preparation for the course.

- *Statistical techniques and end-of-chapter problems are explicitly linked.* After a technique is introduced, students are directed to specific problems for practice and review. The "how-to-do-it" aspects of calculation are immediately and clearly reinforced.
- *End-of-chapter problems are organized progressively.* Simpler problems with small data sets are presented first. Often, explicit instructions or hints accompany the first several problems in a set. The problems gradually become more challenging and require more decision making by the student (e.g., choosing the most appropriate statistic for a certain situation). Thus, each problem set develops problem-solving abilities gradually and progressively.
- *Computer applications.* This text integrates SPSS, the leading social science statistics package, to help students take advantage of the power of the computer. Appendix F provides an introduction to SPSS, and demonstrations are integrated into the chapters. SPSS-based problems are included at the end of chapters, and research projects using SPSS are presented in the "You Are the Researcher" feature.
- *Realistic, up-to-date data.* The databases for computer applications in the text include a shortened version of the 2012 General Social Survey, a data set that includes census and crime data for the 50 states, and a data set that includes demographic data for 99 nations. These databases will give students the opportunity to practice their statistical skills on "real-life" data. All databases are described in Appendix G.

Additional Course Design Resources

- *Online PowerPoint® Slides.* A revised series of PowerPoint slides allows instructors to deliver class lectures and presentations discussing chapter-by-chapter content.
- *Online Instructor's Manual/Testbank.* The Instructor's Manual includes chapter summaries, a test item file of multiple-choice questions, answers to even-numbered computational problems, and step-by-step solutions to selected problems. In addition, the Instructor's Manual includes cumulative exercises (with answers) that can be used for testing purposes. To access these instructor resources, please log in to your account at **http://login.cengage.com**.
- *Aplia*™ is an online interactive learning solution that can be assigned as part of the course. Aplia integrates a variety of media and tools such as video, tutorials, practice tests, and an interactive e-book, and provides students with detailed, immediate feedback on every question. For more information about how to use Aplia in your course, please work with your local Cengage Learning Consultant.

Changes to the Tenth Edition

The following are the most important changes to this edition:

- SPSS has been moved to a more central place in the text:
    - Almost all chapters have new sections ("Using SPSS") that illustrate how to produce the statistics covered in the chapter.

  - SPSS problems have been added to the end-of-chapter problems throughout the text. In some chapters (e.g., Chapters 13 and 15), the SPSS problems replace problems using hand calculators.
  - For statistics that require complex computation—such as Pearson's *r* (Chapter 13) and partial correlation, multiple correlation, and regression (Chapter 15)—explanations and examples are now SPSS-based.
- The data sets used in the text have been expanded and updated. The data sets are used throughout the text in the new "Using SPSS" sections, in the end-of-chapter problems, and in the "You Are the Researcher" projects at the end of most chapters. The data sets are available for downloading at the Web site for this text: **www.cengagebrain.com;** they include:
  - A General Social Survey (GSS) data set (*GSS2012.sav*), which has been updated to 2012.
  - A data set that includes census and crime data on the 50 states (*States.sav*).
  - A data set that includes demographic data for 99 nations (*Intl-POP.sav*).
  - A fourth data set (*CrimeTrends84-10.sav*) is used only for the graphing exercises in Chapter 2.
- Former Chapters 12 and 13 have been combined into a single chapter (Chapter 12, entitled "Bivariate Association for Nominal- and Ordinal-Level Variables"). This new chapter de-emphasizes phi and the mechanics of computation for gamma but still fully treats the analysis of association for variables organized in bivariate tables.
- To accommodate the greater attention given to SPSS and other new material, a number of statistics and techniques have been deleted, including the index of qualitative variation (IQV, formerly in Chapter 4), the computation of means, medians, and the standard deviation for grouped data, and the section on testing gamma and Spearman's rho for significance.
- Chapter 2 has been reorganized and now begins with frequency distributions.
- Boxplots have been added to Chapter 4.
- All chapters now begin with a "Using Statistics" box that cites examples of how the statistics presented in the chapter can be applied to social research and to everyday life.
- Most of the "Statistics in Everyday Life" boxes have been updated or changed.
- The "Reading Statistics" inserts have been updated.
- The data sets used for examples, in the boxed features, and in the end-of-chapter problems have been updated.
- Titles have been added to all boxed features to clarify content and purpose.
- Icons have been added to all boxed inserts to highlight their purpose and content.

Numerous other changes have been made throughout the text, most of them minor. All are intended to clarify explanations and make the material more accessible to students. As with previous editions, my goal is to offer a comprehensive, flexible, and student-oriented book that will provide a challenging first exposure to social statistics.

Acknowledgments

This text has been in development for 30 years. An enormous number of people have made contributions, both great and small, to this project and, at the risk of inadvertently omitting someone, I am bound to at least attempt to acknowledge my many debts.

Much of whatever integrity and quality this book has is a direct result of the very thorough (and often highly critical) reviews that have been conducted over the years. I am consistently impressed by the sensitivity of my colleagues to the needs of the students, and, for their assistance in preparing this edition, I would like to thank the following reviewers: Dawn Baunach, Georgia State University; Mary Bernstein, University of Connecticut; Miyuki Fukushima, Cleveland State University; Michael Gillespie, Eastern Illinois University; Melodie Hallett, San Diego State University; Bennett Judkins, Lee University; Shannon Monnat, University of Nevada, Las Vegas; and John Shandra, Stony Brook University. Whatever failings are contained in the text are, of course, my responsibility and are probably the result of my occasional decisions not to follow the advice of my colleagues.

I would like to thank the instructors who made statistics understandable to me (Professors Satoshi Ito, Noelie Herzog, and Ed Erikson) and all of my colleagues at Christopher Newport University for their support and encouragement (especially Professors F. Samuel Bauer, Stephanie Byrd, Cheryl Chambers, Robert Durel, Marcus Griffin, Mai Lan Gustafsson, Kai Heiddemann, Ruth Kernodle, Michael Lewis, Marion Manton, Eileen O'Brien, Lea Pellet, Eduardo Perez, Virginia Purtle, Andrea Timmer, and Linda Waldron). I owe a special debt of gratitude to Professor Roy Barnes of the University of Michigan–Flint for his help with the "Interpreting Statistics" feature. Also, I thank all of my students for their patience and thoughtful feedback, and I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint Appendices B, C, and D from their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).

Finally, I want to acknowledge the support of my family and rededicate this work to them. I have the extreme good fortune to be a member of an extended family that is remarkable in many ways and that continues to increase in size. Although I cannot list everyone, I would especially like to thank the older generation (my mother, Alice T. Healey), my wife Patricia A. Healey, the next generation (my sons, Kevin and Christopher, my daughters-in-law, Jennifer and Jessica, my step-son Christopher Schroen, and step-daughters Kate Cowell and her husband Matt and Jennifer Schroen), and the youngest generation (Benjamin, Caroline, and Isabelle Healey and Abigail Cowell).

Prologue: Basic Mathematics Review

You will probably be relieved to hear that this text, your first exposure to statistics for social science research, does not stress computation per se. While you will encounter many numbers to work with and numerous formulas to solve, the major emphasis will be on understanding the role of statistics in research and the logic we use to answer research questions empirically. You will also find that, in this text, the example problems and many of the homework problems have been intentionally simplified so that the computations will not unduly impede the task of understanding the statistics themselves.

On the other hand, you may regret to learn that there is, inevitably, some arithmetic that you simply cannot avoid if you want to master this material. It is likely that some of you haven't had any math in a long time, others have convinced themselves that they just cannot do math under any circumstances, and still others are just rusty and out of practice. All of you will find that mathematical operations that might seem complex and intimidating at first glance can be broken down into simple steps. If you have forgotten how to cope with some of these steps or are unfamiliar with these operations, this prologue is designed to ease you into the skills you will need to do all of the computations in this textbook. Also, you can use this section for review whenever you feel uncomfortable with the mathematics in the chapters to come.

Calculators and Computers

A calculator is a virtual necessity for this text. Even the simplest, least expensive model will save you time and effort and is definitely worth the investment. However, I recommend that you consider investing in a more sophisticated calculator with memories and preprogrammed functions, especially the statistical models that can compute means and standard deviations automatically. Calculators with these capabilities are available for around $20.00 to $30.00 and will almost certainly be worth the small effort it will take to learn to use them.

In the same vein, there are several computerized statistical packages (or **statpaks**) commonly available on college campuses that can further enhance your statistical and research capabilities. The most widely used of these is the Statistical Package for the Social Sciences (**SPSS**). Statistical packages like SPSS are many times more powerful than even the most sophisticated handheld calculators, and it will be well worth your time to learn how to use them because they will eventually save you time and effort.

SPSS is introduced in Appendix F of this text and is integrated into almost all the chapters. There are demonstrations that show you, step by step, how to use the program to generate the statistics covered in the chapter and end-of-chapter problems that require you to apply the program. Furthermore, the "You Are the Researcher" feature at the end of most chapters gives you the opportunity to use SPSS in some simplified social research projects.

There are many other programs that can help you calculate statistics with a minimum of effort and time. Even spreadsheet programs such as Microsoft® Excel, which is included in many versions of Microsoft Office, have some statistical capabilities. You should be aware that all of these programs (other than the simplest calculators) will require some effort to learn, but the rewards will be worth the effort.

In summary, you should find a way at the beginning of this course—with a calculator, a statpak, or both—to minimize the tedium of mere computing. This will permit you to devote maximum effort to the truly important goal of increasing your understanding of the meaning of statistics in particular and social research in general.

Variables and Symbols

Statistics are a set of techniques by which we can describe, analyze, and manipulate variables. A **variable** is a trait that can change value from case to case or from time to time. Examples of variables include height, weight, level of prejudice, and political party preference. The possible values or scores associated with a given variable might be numerous (for example, income) or relatively few (for example, gender). I will often use symbols, usually the letter X, to refer to variables in general or to a specific variable.

Sometimes we will need to refer to a specific value or set of values of a variable. This is usually done with the aid of subscripts. So the symbol X_1 (read "X-sub-one") would refer to the first score in a set of scores, X_2 ("X-sub-two") to the second score, and so forth. Also, we will use the subscript i to refer to all the scores in a set. Thus, the symbol X_i ("X-sub-i") refers to all of the scores associated with a given variable (for example, the test grades of a particular class).

Operations

You are all familiar with the four basic mathematical operations of addition, subtraction, multiplication, and division and the standard symbols (+, −, ×, ÷) used to denote them. I should remind you that multiplication and division can be symbolized in a variety of ways. For example, the operation of multiplying some number a by some number b may be symbolized in (at least) six different ways:

$$\mathrm{a} \times \mathrm{b}$$
$$a \cdot b$$
$$a * b$$
$$ab$$
$$a(b)$$
$$(a)(b)$$

In this text, we will commonly use the “adjacent symbols” format (that is, *ab*), the conventional times sign ($\times$), or adjacent parentheses to indicate multiplication. On most calculators and computers, the asterisk (*) is the symbol for multiplication.

The operation of division can also be expressed in several different ways. In this text, in addition to the standard symbol for division, we will use either of these two methods:

$$a/b \text{ or } \frac{a}{b}$$

Several formulas require us to find the square of a number. To do this, multiply the number by itself. This operation is symbolized as X^2 (read “X squared”), which is the same thing as $(X)(X)$. If X has a value of 4, then

$$X^2 = (X)(X) = (4)(4) = 16$$

or we could say that “4 squared is 16.”

The square root of a number is the value that, when multiplied by itself, results in the original number. So the square root of 16 is 4 because (4)(4) is 16. The operation of finding the square root of a number is symbolized as

$$\sqrt{X}$$

A final operation with which you should be familiar is summation, or the addition of the scores associated with a particular variable. When a formula requires the addition of a series of scores, this operation is usually symbolized as ΣX_i. “Σ” is uppercase Greek letter sigma and stands for “the summation of.” So the combination of symbols ΣX_i means “the summation of all the scores” and directs us to add all the scores for that variable. If four people had family sizes of 2, 4, 5, and 7, then the summation of these four scores for this variable could be symbolized as

$$\Sigma X_i = 2 + 4 + 5 + 7 = 18$$

The symbol Σ is an operator, just like the $+$ and $\times$ signs. It directs us to add all of the scores on the variable indicated by the X symbol.

There are two other common uses of the summation sign, and, unfortunately, the symbols denoting these uses are not, at first glance, sharply different from each other or from the symbol used earlier. Some careful attention to these various meanings should minimize the confusion.

The first set of symbols is ΣX^2, which means “the sum of the squared scores.” This quantity is found by *first* squaring each of the scores and *then* adding the squared scores together. A second common set of symbols will be $(\Sigma X_i)^2$, which means “the sum of the scores, squared.” This quantity is found by *first* summing the scores and *then* squaring the total.

These distinctions might be confusing at first, so let’s use an example to help clarify the situation. Suppose we had a set of three scores: 10, 12, and 13. So

$$X_i = 10, 12, 13$$

The sum of these scores is

$$\Sigma X_i = 10 + 12 + 13 = 35$$

The sum of the squared scores would be

$$\Sigma X^2 = (10)^2 + (12)^2 + (13)^2 = 100 + 144 + 169 = 413$$

Take careful note of the order of operations here. First the scores are squared one at a time, and then the squared scores are added. This is a completely different operation from squaring the sum of the scores:

$$(\Sigma X_i)^2 = (10 + 12 + 13)^2 = (35)^2 = 1225$$

To find this quantity, first the scores are summed and then the total of all the scores is squared. The squared sum of the scores (1225) is *not* the same as the sum of the squared scores (413).

In summary, the operations associated with each set of symbols can be summarized as follows:

Symbol	Operations
ΣX_i	Add the scores
ΣX_i^2	First square the scores and then add the squared scores
$(\Sigma X_i)^2$	First add the scores and then square the total

Operations with Negative Numbers

A number can be either positive (if it is preceded by a + sign or by no sign at all) or negative (if it is preceded by a − sign). Positive numbers are greater than 0, and negative numbers are less than 0. It is very important to keep track of signs because they will affect the outcome of virtually every mathematical operation. This section briefly summarizes the relevant rules for dealing with negative numbers.

First, adding a negative number is the same as subtraction. For example,

$$3 + (-1) = 3 - 1 = 2$$

Second, subtraction changes the sign of a negative number:

$$3 - (-1) = 3 + 1 = 4$$

Note the importance of keeping track of signs here. If you neglected to change the sign of the negative number in the second expression, you would get the wrong answer.

For multiplication and division, you need to be aware of various combinations of negative and positive numbers. Ignoring the case of all positive numbers, this leaves several possible combinations. A negative number times a positive number results in a negative value:

$$(-3)(4) = -12$$
$$(3)(-4) = -12$$

A negative number multiplied by a negative number is always positive:

$$(-3)(-4) = 12$$

Division follows the same patterns. If there is a negative number in the calculations, the answer will be negative. If both numbers are negative, the answer will be positive. So

$$\frac{-4}{2} = -2$$

and

$$\frac{4}{-2} = -2$$

but

$$\frac{-4}{-2} = 2$$

Negative numbers do not have square roots, because multiplying a number by itself cannot result in a negative value. Squaring a negative number always results in a positive value (see the multiplication rules earlier).

Accuracy and Rounding Off

A possible source of confusion in computation involves accuracy and rounding off. People work at different levels of precision and, for this reason alone, may arrive at different answers to problems. This is important because our answers can be at least slightly different if you work at one level of precision and I (or your instructor or your study partner) work at another. You may sometimes think you've gotten the wrong answer when all you've really done is round off at a different place in the calculations or in a different way.

There are two issues here: when to round off and how to round off. My practice is to work in as much accuracy as my calculator or statistics package will allow and then round off to two places of accuracy (two places beyond, or to the right of, the decimal point) only at the very end. If a set of calculations is lengthy and requires the reporting of intermediate sums or subtotals, I will round the subtotals off to two places as I go.

In terms of how to round off, begin by looking at the digit immediately to the right of the last digit you want to retain. If you want to round off to 100ths (two places beyond the decimal point), look at the digit in the 1000ths place (three places beyond the decimal point). If that digit is 5 or more, round up. For example, 23.346 would round off to 23.35. If the digit to the right is less than 5, round down. So, 23.343 would become 23.34.

Let's look at an additional example of how to follow these rules of rounding. If you are calculating the mean value of a set of test scores and your calculator shows a final value of 83.459067 and you want to round off to two places, look at the digit three places beyond the decimal point. In this case the value is 9 (greater than 5), so we would round the second digit beyond the decimal point up and report the mean as 83.46. If the value had been 83.453067, we would have reported our final answer as 83.45.

Formulas, Complex Operations, and the Order of Operations

A mathematical formula is a set of directions, stated in general symbols, for calculating a particular statistic. To "solve a formula," you replace the symbols with the proper values and then perform a series of calculations. Even the most complex formula can be simplified by breaking the operations down into smaller steps.

Working through the steps requires some knowledge of general procedure and the rules of precedence of mathematical operations. This is because the order in which you perform calculations may affect your final answer. Consider the following expression:

$$2 + 3(4)$$

If you add first, you will evaluate the expression as

$$5(4) = 20$$

but if you multiply first, the expression becomes

$$2 + 12 = 14$$

Obviously, it is crucial to complete the steps of a calculation in the correct order.

The basic rules of precedence are to find all squares and square roots first, then do all multiplication and division, and finally complete all addition and subtraction. So the expression

$$8 + 2 \times 2^2/2$$

would be evaluated as

$$8 + 2 \times \frac{4}{2} = 8 + \frac{8}{2} = 8 + 4 = 12$$

The rules of precedence may be overridden by parentheses. Solve all expressions within parentheses before applying the rules stated earlier. For most formulas in this text, the order of calculations will be controlled by the parentheses. Consider the following expression:

$$(8 + 2) - \frac{4}{(5 - 1)}$$

Resolving the parenthetical expressions first, we would have

$$(8 + 2) - \frac{4}{(5 - 1)} = (10) - \frac{4}{4} = 10 - 1 = 9$$

A final operation you will encounter in some formulas in this text involves denominators of fractions that themselves contain fractions. In this situation, solve the fraction in the denominator first and then complete the division. For example,

$$\frac{15 - 9}{6/2}$$

would become

$$\frac{15 - 9}{6/2} = \frac{6}{3} = 2$$

When you are confronted with complex expressions such as these, don't be intimidated. If you're patient with yourself and work through them step by step, beginning with the parenthetical expressions, even the most imposing formulas can be managed.

Exercises

You can use the following problems as a self-test on the material presented in this review. If you can handle these problems, you're ready to do all of the arithmetic in this text. If you have difficulty with any of these problems, please review the appropriate section of this prologue. You might also want to use this section as an opportunity to become more familiar with your calculator. The answers are given immediately following these exercises, along with commentary and some reminders.

1. Complete each of the following:
   a. $17 \times 3 =$
   b. $17(3) =$
   c. $(17)(3) =$
   d. $17/3 =$
   e. $(42)^2 =$
   f. $\sqrt{113} =$

2. For the set of scores (X_i) of 50, 55, 60, 65, and 70, evaluate each of the following expressions:
   $\Sigma X_i =$
   $\Sigma X_i^2 =$
   $(\Sigma X_i)^2 =$

3. Complete each of the following:
   a. $17 + (-3) + (4) + (-2) =$
   b. $(-27)(54) =$
   c. $(-14)(-100) =$
   d. $-34/(-2) =$
   e. $322/(-11) =$
   f. $\sqrt{-2} =$
   g. $(-17)^2 =$

4. Round off each of the following to two places beyond the decimal point:
   a. 17.17532
   b. 43.119
   c. 1076.77337
   d. 32.4641152301
   e. 32.4751152301

5. Evaluate each of the following:
- **a.** $(3 + 7)/10 =$
- **b.** $3 + 7/10 =$
- **c.** $\dfrac{(4 - 3) + (7 + 2)(3)}{(4 + 5)(10)} =$
- **d.** $\dfrac{22 + 44}{15/3} =$

Answers to Exercises

1. **a.** 51 **b.** 51 **c.** 51
(The obvious purpose of these first three problems is to remind you that there are several different ways of expressing multiplication.)
d. 5.67 (Note the rounding off.) **e.** 1764 **f.** 10.63

2. The first expression is "the sum of the scores," so this operation would be

$$\Sigma X_i = 50 + 55 + 60 + 65 + 70 = 300$$

The second expression is the "sum of the squared scores." So

$$\Sigma X_i^2 = (50)^2 + (55)^2 + (60)^2 + (65)^2 + (70)^2$$
$$\Sigma X_i^2 = 2500 + 3025 + 3600 + 4225 + 4900$$
$$\Sigma X_i^2 = 18{,}250$$

The third expression is "the sum of the scores, squared":

$$(\Sigma X_i)^2 = (50 + 55 + 60 + 65 + 70)^2$$
$$(\Sigma X_i)^2 = (300)^2$$
$$(\Sigma X_i)^2 = 90{,}000$$

Remember that ΣX_i^2 and $(\Sigma X_i)^2$ are two completely different expressions with very different values.

3. **a.** 16
b. −1458
c. 1400
d. 17
e. −29.27
f. Your calculator probably gave you some sort of error message for this problem, because negative numbers do not have square roots.
g. 289

4. **a.** 17.18 **b.** 43.12 **c.** 1076.77
d. 32.46 **e.** 32.48

5. **a.** 1 **b.** 3.7 (Note the importance of parentheses.) **c.** 0.31
d. 13.2

1 Introduction

LEARNING OBJECTIVES

By the end of this chapter, you will be able to

1. Describe the limited but crucial role of statistics in social research.
2. Distinguish among three applications of statistics and identify situations in which each is appropriate.
3. Distinguish between discrete and continuous variables and cite examples of each.
4. Identify and describe three levels of measurement and cite examples of variables from each.

USING STATISTICS

One of the most important themes of this text is that statistics are useful tools for analyzing and understanding information and for communicating our conclusions to others. To stress this theme, each chapter will begin with a list of situations in which statistics can be (and should be) applied to good advantage. In this introductory chapter, we will focus on the most general examples, but in the rest of the text, this section will highlight the usefulness of the specific statistics presented in that chapter.

Statistics can be used to

- Demonstrate the connection between smoking and cancer
- Measure political preferences, including the popularity of specific candidates for office
- Track attitudes about gay marriage, abortion, and other controversial issues over time
- Compare the cost of living (housing prices and rents, the cost of groceries and gas, health care, and so forth) between different localities (cities, states, and even nations)

Why Study Statistics?

Students sometimes wonder about the value of studying statistics. What, after all, do numbers and formulas have to do with understanding people and society? The value of statistics will become clear as we move from chapter to chapter but, for now, we can demonstrate the importance of statistics by considering **research** or disciplined and careful inquiries. Scientists conduct research to answer questions, examine ideas, and test theories. Research can take numerous forms, and statistics are relevant for **quantitative research** projects, or projects that collect information in the form of numbers or **data**. **Statistics** are mathematical techniques used by social scientists to analyze data in order to answer questions and test theories.

What is so important about learning how to manipulate data? On one hand, some of the most important and enlightening works in the social sciences do not utilize statistics at all. There is nothing magical about data and statistics. The mere presence of numbers guarantees nothing about the quality of a scientific inquiry.

On the other hand, data can be the most trustworthy information available to the researcher and, consequently, deserve special attention. Data that have been carefully collected and thoughtfully analyzed can be the strongest, most objective foundations for building theory and enhancing our understanding of the social world.

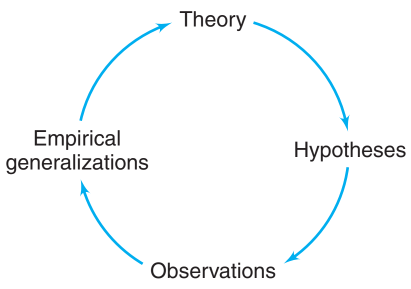
Let us be very clear about one point: It is never enough merely to gather data (or, for that matter, any kind of information). Even the most carefully collected numerical information does not and cannot speak for itself. The researcher must be able to use statistics effectively to organize, evaluate, and analyze the data. Without a good understanding of statistical analysis, the researcher will be unable to make sense of the data. Without an appropriate application of statistics, data are useless.

Statistics are an indispensable tool for the social sciences. They provide the scientist with some of the most useful techniques for evaluating ideas and testing theory. The next section describes the relationships among theory, research, and statistics in more detail.

The Role of Statistics in Scientific Inquiry

Figure 1.1 graphically represents the role of statistics in the research process. The diagram is based on the thinking of Walter Wallace and illustrates a useful conception of how any scientific enterprise grows and develops. One point the diagram makes is that scientific theory and research continually shape each other.

FIGURE 1.1 The Wheel of Science



Source: Adapted from Walter Wallace, *The Logic of Science in Sociology* (Chicago: Aldine-Atherton, 1971).

Statistics are one of the most important means by which research and theory interact. Let's take a closer look at the process.

A JOURNEY THROUGH THE SCIENTIFIC PROCESS

Because Figure 1.1 is circular, with no beginning or end, we could start our journey at any point. For the sake of convenience, let's begin at the top, with theory, and follow the arrows around the circle. A **theory** is an explanation of the relationships among phenomena. People naturally (and endlessly) wonder about problems in society (such as prejudice, poverty, child abuse, and serial murders), and they develop explanations ("low levels of education cause prejudice") in their attempts to understand. Unlike our everyday, informal explanations, scientific theory is subject to a rigorous testing process. Let's take prejudice as an example to illustrate how the research process works.

Theory. Why are some people prejudiced against other groups? One possible answer to this question is provided by a theory of prejudice stated over 50 years ago by social psychologist Gordon Allport and tested on a number of occasions since that time.[1]

The theory states that prejudice will decrease in situations in which members of different groups have equal status and cooperate with each other to work toward mutual goals. The more equal and cooperative the contact, the more likely people will see each other as individuals and not as representatives of a particular group. For example, we might predict that members of a racially mixed athletic team who work together to achieve victory will experience a decline in prejudice. On the other hand, when different groups compete for jobs, housing, or other resources, prejudice will tend to increase. Allport's theory is not a complete explanation of prejudice, of course, but it will serve to illustrate a sociological theory.

Variables and Causation. Note that Allport's theory is stated in terms of causal relationships between two variables. A **variable** is any trait that can change values from case to case; examples include gender, age, income, and political party affiliation. A theory may identify some variables as causes and others as effects or results. In the language of science, the causes are **independent variables** and the effects or results are **dependent variables**. In our theory, equal-status contact would be the independent variable (or the cause) and prejudice would be the dependent variable (the result or effect). In other words, the theory argues that prejudice *depends on* (or is caused by) the extent to which a person participates in equal-status, cooperative contacts with members of other groups.

[1] Allport, Gordon, 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley. This theory is often called "the contact hypothesis." For recent attempts to test this theory, see: McLaren, Lauren. 2003. "Anti-Immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants." *Social Forces*, 81:909–937; Pettigrew, Thomas. 1997. "Generalized Intergroup Contact Effects on Prejudice." *Personality and Social Psychology Bulletin*, 23:173–185; and Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. 2011. "Recent advances in intergroup contact theory." International Journal of Intercultural Relations, 35: 271–280.

Diagrams can be a useful way of representing the relationships between variables:

$$\text{Equal-status contact} \rightarrow \text{Prejudice}$$
$$\text{Independent variable} \rightarrow \text{Dependent variable}$$
$$X \rightarrow Y$$

The arrow represents the direction of the causal relationship, and X and Y are general symbols for the independent and dependent variables, respectively.

Hypotheses. So far, we have a theory of prejudice and independent and dependent variables. Is the theory true or false? To find out, we need to compare our theory with the facts; we need to do some research. Our next steps would be to define our terms and ideas. One problem we often face in research is that theories are complex and abstract, and we need to be very specific to conduct a valid test. Often, we do this by deriving a hypothesis from the theory: A **hypothesis** is a statement about the relationship between variables that is specific and exact.

For example, if we wished to test Allport's theory, we would have to say exactly what we mean by prejudice and we would need to describe "equal-status, cooperative contact" in detail. We would also review the research literature to help develop and clarify our definitions and our understanding of these concepts.

As our hypothesis takes shape, we begin the next step of the research process, during which we decide exactly how we will gather data. We must decide how cases will be selected and tested, how variables will be measured, and a host of related matters. Ultimately, these plans will lead to the observation phase (the bottom of the wheel of science), where we actually measure social reality. Before we can do this, we must have a very clear idea of what we are looking for and a well-defined strategy for conducting the search.

Making Observations and Using Statistics. To test Allport's theory of prejudice, we might begin with people from different racial or ethnic groups. We might place some subjects in cooperative situations and others in competitive situations. We would need to measure levels of prejudice before and after each type of contact. We might do this by administering a survey that asks subjects to agree or disagree with statements such as "Greater efforts must be made to racially integrate the public school system." Our goal would be to see whether the people exposed to the cooperative-contact situations actually become less prejudiced.

Now, finally, we come to the use of statistics in our research. During the observation phase, we will collect a great deal of numerical information or data. If we have a sample of 100 people, we will have 200 completed surveys measuring prejudice: 100 completed before the contact situation and 100 filled out afterward. Try to imagine dealing with 200 completed surveys. If we ask each respondent just five questions to measure their prejudice, we will have a total of 1000 separate pieces of information to deal with. What will we do? We'll have to organize and analyze this information, and statistics will become very helpful at this point. They will supply us with many ideas about "what to do" with the data. We will begin to look at some of the options in the next chapter; for now, let me stress two points about statistics.

First, statistics are crucial. Simply put, without statistics, quantitative research is impossible. We need statistics to analyze data and to shape and refine our theories of the social world.

Second, and somewhat paradoxically, the role of statistics is limited. As Figure 1.1 makes clear, scientific research proceeds through several mutually interdependent stages. Statistics become directly relevant only at the end of the observation stage. Before any statistical analysis can be legitimately applied, however, earlier phases must have been successfully completed. If the researcher has asked poorly conceived questions or has made serious errors of design or method, then even the most sophisticated statistical analysis is valueless. As useful as they can be, statistics cannot substitute for rigorous conceptualization, detailed and careful planning, or creative use of theory. Statistics cannot salvage a poorly conceived or designed research project. They cannot make sense out of garbage.

On the other hand, inappropriate statistical applications can limit the usefulness of an otherwise carefully done project. Only by successfully completing *all* phases of the process can a quantitative research project hope to contribute to understanding. A reasonable knowledge of the uses and limitations of statistics is as essential to the education of the social scientist as is training in theory and methodology.

Empirical Generalizations. Our statistical analysis would focus on assessing our theory, but we would also look for other trends in the data. For example, if we found that equal-status, cooperative contact reduces prejudice in general, we might go on to ask whether the pattern applies to males as well as to females and to the well educated as well as to the poorly educated. As we probed the data, we might begin to develop some generalizations based on the empirical patterns we observe. For example, what if we found that contact reduced prejudice for younger respondents but not for older respondents? Could it be that younger people are less "set in their ways" and have attitudes and feelings that are more open to change? As we developed tentative explanations, we would begin to revise or elaborate the theory that guides the research project.

STATISTICS
IN EVERYDAY LIFE

Push Polls

Political campaigns sometimes use "push polls" to sway public opinion. These polls are designed to influence opinions, sometimes by starting or circulating rumors and innuendo. For example, they may try to brand an opponent as untrustworthy by asking questions like "Would you still support the candidate if you found out that he is an alcoholic?" Even when completely fabricated, such a question can create negative associations in the minds of voters. Statistics may be used to "analyze" data gathered by push polls (or by marketing campaigns that use similar techniques), but the results have little or no scientific validity.

A New Theory? If we changed our theory because of our empirical generalizations, a new research project would probably be needed to test the revised theory, and the wheel of science would begin to turn again. We (or perhaps other researchers) would go through the entire process once again with this new—and, hopefully, improved—theory. This second project might result in further revisions that would require still more research, and the wheel of science would continue to turn as long as scientists were able to suggest revisions or develop new insights. Every time the wheel turned, our understanding of the phenomenon under consideration would (hopefully) improve.

Summary: The Dialog Between Theory and Research. This description of the research process does not include white-coated, clipboard-carrying scientists who, in a blinding flash of inspiration, discover some fundamental truth about reality and shout "Eureka!" The truth is that, in the normal course of science, we rarely are in a position to declare a theory true or false. Rather, evidence will gradually accumulate over time, and ultimate judgments of truth will be the result of many years of hard work, research, and debate.

Let's briefly review our imaginary research project. We began with a theory and examined the various stages of the research project that would be needed to test that theory. We wound up back at the top of the wheel, ready to begin a new project guided by a revised theory. We saw how theory can motivate research and how our observations might cause us to revise theory and, thus, motivate a new research project: Theory stimulates research and research shapes theory. This constant interaction is the lifeblood of science and the key to enhancing our understanding of the social world.

The dialog between theory and research occurs at many levels and in multiple forms. Statistics are one of the most important links between these two realms. Statistics permit us to analyze data, to identify and probe trends and relationships, to develop generalizations, and to revise and improve our theories. As you will see throughout this text, statistics are limited in many ways. They are also an indispensable part of the research enterprise, an essential tool for conducting research and shaping theory. *(For practice in describing the relationship between theory and research and the role of statistics in research, see problems 1.1 and 1.2.)*

The Goals of This Text

Clearly, statistics are a crucial part of social science research and every social scientist needs some training in statistical analysis. In this section, we address how much training is necessary and the purpose of that training.

First, this textbook takes the view that statistics are tools: useful, but not ends in themselves. Thus, we will not take a "mathematical" approach to the subject, although we will cover enough material so that you can develop a basic understanding of why statistics do what they do. Instead, statistics will be presented as tools that can be used to answer important questions, and our focus will be on how these techniques are applied in the social sciences.

Second, you will soon become involved in advanced coursework in your major fields of study, and you will find that much of the professional literature assumes at least basic statistical literacy. Furthermore, after graduation, many of you will find yourselves in positions—either in a career or in graduate school—where some understanding of statistics will be very helpful or perhaps even required. Very few of you will become statisticians per se, but you must have a grasp of statistics in order to read and critically appreciate the research literature of your discipline. As a student in the social sciences and in many careers related to the social sciences, you simply cannot realize your full potential without a background in statistics.

Within these constraints, this textbook is an introduction to statistics as they are utilized in the social sciences. The general goal of the text is to develop an appreciation—a healthy respect—for statistics and their place in the research process. You should emerge from this experience with the ability to use statistics intelligently and to know when other people have done so—and when they have not! You should be familiar with the advantages and limitations of the more commonly used statistical techniques, and you should know which techniques are appropriate for a given purpose. Finally, you should develop sufficient statistical and computational skills and enough experience in the interpretation of statistics to be able to carry out some elementary forms of data analysis by yourself.

Descriptive and Inferential Statistics

As noted earlier, statistics are tools used to analyze data and answer research questions. Two general classes of statistical techniques, introduced in this section, are available to accomplish this task.

DESCRIPTIVE STATISTICS

The first class of techniques, known as **descriptive statistics**, is relevant in several different situations:

1. When a researcher needs to summarize or describe the distribution of a single variable. These statistics are called *univariate* ("one variable") descriptive statistics.
2. When the researcher wishes to describe the relationship between two or more variables. These statistics are called *bivariate* ("two variable") or *multivariate* (more than two variable) descriptive statistics.

Univariate Descriptive Statistics. Many of these statistics are probably familiar to you. For example, percentages, averages, and graphs can all be used to describe single variables.

To illustrate the usefulness of univariate descriptive statistics, consider the following problem: Suppose you wanted to summarize the income distribution of a community of 10,000 families. How would you do it? Obviously, you couldn't simply list all the incomes and let it go at that. You would want to use some summary measures of the overall distribution—perhaps a graph, an average, or the

STATISTICS
IN EVERYDAY LIFE

Using Descriptive Statistics

American society is increasingly connected to the Internet. In 2011, about 78% of all American adults used the Internet at least occasionally—a dramatic increase from 53% in 2000. However, connectedness is very dependent on social class: 96% of the most affluent Americans used the Internet vs. only 63% of the least affluent. What are the implications of this pattern? Are the less affluent being left behind? Will their lower use of this increasingly essential resource affect the education and job prospects of their children and thus perpetuate their lower income over the generations? What additional information would you need to answer these questions?

Source: U.S. Bureau of the Census. 2012. *Statistical Abstract of the United States: 2012*. Available at http://www.census.gov/prod/2011pubs/12statab/infocomm.pdf.

percentage of incomes that are low, moderate, or high. Whatever method you choose, its function is to reduce these thousands of items of information into a few clear, concise, and easily understood numbers. The process of using a few numbers to summarize many, called **data reduction**, is the basic goal of univariate descriptive statistics. Part I of this text is devoted to these statistics.

Bivariate and Multivariate Descriptive Statistics. The second type of descriptive statistics is designed to help us understand the relationship between two or more variables. These statistics, called **measures of association**, allow us to quantify the strength and direction of a relationship. We can use these statistics to investigate two matters of central theoretical and practical importance to any science: causation and prediction. Measures of association help us disentangle the connections between variables and trace the ways in which some variables might affect others. We can also use them to predict scores on one variable from scores on another.

For example, suppose you were interested in the relationship between two variables—amount of study time and grades—and had gathered data from a group of college students. By calculating the appropriate measure of association, you could determine the strength of the relationship and its direction. Suppose you found a strong, positive relationship. This would indicate that "study time" and "grade" were closely related (strength of the relationship) and that as one increased in value, the other also increased (direction of the relationship). You could make predictions from one variable to the other ("the longer the study time, the higher the grade").

Measures of association can give us valuable information about relationships between variables and help us understand how one variable causes another. One important point to keep in mind about these statistics, however, is that they cannot, by themselves, *prove* that two variables are causally related. Even if a measure of association shows a very strong relationship between study time and grades, we cannot conclude that one variable causes the other. Correlation is not the same thing as causation, and the mere existence of a correlation cannot prove that a causal relationship exists between variables. We will consider bivariate associations or correlations in Part III of this text, and we will cover multivariate analysis in Part IV.

STATISTICS
IN EVERYDAY LIFE

Using Inferential Statistics

In 2012, a sample of 3008 adult Americans was asked about gay marriage. About 47% were in favor of allowing gays and lesbians to marry legally (up from 31% in 2004) and 43% were opposed (down from 60% in 2004). Some say that American society is polarized on the issue of gay marriage. Do these statistics support that view? How?

Source: Pew Research Center. Available at http://www.people-press.org/2012/04/25/more-support-for-gun-rights-gay-marriage-than-in-2008-or-2004/1/.

INFERENTIAL STATISTICS

This second class of statistical techniques becomes relevant when we wish to generalize to a **population**, the total collection of all cases in which the researcher is interested and wishes to understand better. Examples of possible populations would be voters in the United States, all parliamentary democracies, or all unemployed people in Atlanta.

Populations can theoretically range from enormous ("all humanity") to quite small (all sophomores on your campus) but are usually fairly large. Social scientists almost never have the resources or time to test every case in a population, hence the need for **inferential statistics**. This class of techniques involves using information from **samples** (carefully chosen subsets of the population) to make inferences about populations. Because they have fewer cases, samples are cheaper to assemble, and—if the proper techniques are followed—generalizations based on samples can be very accurate representations of populations.

Many of the concepts and procedures involved in inferential statistics may be unfamiliar, but most of us are experienced consumers of inferential statistics—most familiarly, perhaps, in the form of public-opinion polls and election projections. When a public-opinion poll reports that 42% of the American electorate plans to vote for a certain presidential candidate, it is essentially reporting a generalization to a population ("the American electorate"—which numbers over 130 million people) from a carefully drawn sample (usually between 1000 and 3000 respondents). Inferential statistics will occupy our attention in Part II of this book. *(For practice in describing different statistical applications, see problems 1.3 and 1.7.)*

Discrete and Continuous Variables

In the next chapter, you will begin to encounter the many statistics available to the social scientist. One aspect of using statistics that can be puzzling is deciding when to use which statistic. You will learn specific guidelines as you go along, but I will introduce some basic and general guidelines at this point. The first of these concerns discrete and continuous variables; the second, covered in the next section, concerns level of measurement.

DISCRETE VARIABLES

A variable is said to be **discrete** if it has a basic unit of measurement that cannot be subdivided. For example, number of people per household is a discrete