BRUCE M. KING • PATRICK J. ROSOPA • EDWARD W. MINIUM

# STATISTICAL REASONING IN THE BEHAVIORAL SCIENCES

Seventh Edition


WILEY

# Statistical Reasoning in the Behavioral Sciences

# Statistical Reasoning in the Behavioral Sciences

SEVENTH EDITION

**Bruce M. King**

**Patrick J. Rosopa**

**Edward W. Minium**

WILEY

| | |
|---|---|
| Vice President and Director | George Hoffman |
| Editorial Director | Veronica Visentin |
| Assistant Editor | Ethan Lipson |
| Editorial Assistant | Monica Rogers |
| Content Enablement Manager | Judy Howarth |
| Content Management Director | Lisa Wojcik |
| Content Manager | Nichole Urban |
| Senior Content Specialist | Nicole Repasky |
| Production Editor | Kavitha Balashundaram |
| Photo Researcher | Billy Ray |
| Marketing Manager | Christine Kurshner |
| Designer | RDC Publishing |
| Cover Photo Credit | © Sergey Nivens/Alamy Stock Photo |

This book was set in 10/12 pt TimesLTStd by SPi Global and printed and bound by Strategic Content Imaging.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.



ISBN: 978-1-119-37973-7 (PBK)
ISBN: 978-1-119-37982-9 (EVALC)

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

Names: King, Bruce M., author. | Rosopa, Patrick, author. | Minium, Edward W., 1917– author.
Title: Statistical reasoning in the behavioral sciences / by Bruce M. King,
 Patrick J. Rosopa, Edward W. Minium.
Description: Seventh edition. | Hoboken, NJ : John Wiley & Sons, Inc., [2018] |
 Includes index. |
Identifiers: LCCN 2017060694 (print) | LCCN 2018007899 (ebook) | ISBN 9781119379904 (pdf) |
 ISBN 9781119379881 (epub) | ISBN 9781119379737 (pbk.)
Subjects: LCSH: Psychometrics. | Educational statistics.
Classification: LCC BF39 (ebook) | LCC BF39 .K53 2018 (print) |
 DDC 300.72/7—dc23
LC record available at https://lccn.loc.gov/2017060694

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

*To Students Who Want to Learn and Instructors Who Like to Teach*

# Preface

*Statistical Reasoning in the Behavioral Sciences* is emphatically concerned with students' understanding of statistical logic and procedures, as well as an understanding of what happens when the strict requirements of statistical theory meet the circumstances of real-world data. Regardless of one's academic field, it is important that students understand the interplay between statistical questions and answers and research questions and answers so that students see that statistics really has significance for their discipline. Thus, throughout the text, we interweave aspects of experimental design with the statistical topics to which they are closely related. We try to be especially clear on those points in inferential statistics that tend to be stumbling blocks. At the same time, we give full treatment in the first half of the text to descriptive statistics in the belief that once inference is done, one must return to the descriptive measures to assess the meaning of the inquiry; both concepts must be fully understood.

The fourth, fifth, and sixth editions of *Statistical Reasoning* followed the recommendations of the APA Task Force on Statistical Inference (Wilkinson et al., 1999) and we continue to do so in the seventh edition. This includes greater emphasis on effect size, power, and confidence intervals. Beginning with Chapter 14, measures of effect size (both *d* family and *r* family measures) are integrated with the material on hypothesis testing.

The text should serve very well for a one-semester course. Even so, there is more material than can be covered in most one-quarter or one-semester courses. The Instructor's Manual offers suggestions for adapting the book to particular needs.

## Features New to the Seventh Edition

In many introductory courses, instructors require that students learn how to use computers to work statistical problems. Thus, starting with the sixth edition and continuing with the seventh edition, we have added instructions for the use of the statistical software package SPSS. Step-by-step instructions are provided as well as numerous screen shots so that students may easily use the software. We have also made an attempt to update some of the example and end-of-chapter problems in the textbook. In addition, we have also added various photos to the textbook to further engage students and enhance learning. A main organizational change was splitting Chapter 12 from the sixth edition into two chapters for the seventh edition. Specifically, Chapter 12 now focuses primarily on the one sample *z* test while the new Chapter 13 develops understanding of the one sample *t* test. We felt that separating the material into two chapters can make the material easier to teach and easier for students to synthesize. The chapter on chi-square still appears after the chapters on ANOVA in order not to interrupt the presentation of statistical methods used with quantitative variables (from the single-sample *t* test through ANOVA).

Although we demonstrate how to use a powerful, user-friendly software package, it is important to highlight that we do not sacrifice any of the material dealing with the conceptual understanding of a statistical procedure. For example, after thoroughly describing the logic behind a procedure, we complete a problem step-by-step using raw data and formulas, followed by an interpretation of the results. Only after adequate conceptual coverage and a complete example do we demonstrate how to use SPSS (i.e., entering data, navigating menus, and using dialog boxes) to obtain the same results in seconds. For instructors who prefer not to cover SPSS, these sections can be skipped without disrupting the continuity of a chapter.

## Pedagogy

*Statistical Reasoning in the Behavioral Sciences* includes the following pedagogical features:

- learning objectives at the beginning of each chapter,
- a summary at the end of each chapter to reinforce what has been presented,
- key terms and concepts that are boldfaced at their first occurrence in each chapter and listed again for review at the end of each chapter,
- the presentation of definitions, symbols, and equations in the margins of the text so that the student may easily identify important new material,
- special boxed sections called Points of Controversy that introduce students to topical and/or current issues to let them see that statistics is an evolving, vital area of inquiry, and
- review of basic mathematical propositions in an appendix.

Some texts present chapters on correlation and regression after the chapters on hypothesis testing. However, we firmly believe that the material on correlation should come first so that we may more fully develop the material on dependent-groups designs and the $r$ family of measures of effect size.

Finally, many students who enroll in introductory statistics do so with great anxiety. Many statistics textbooks can be written in a very formal and stodgy style and some lack any visual material other than graphs of frequency distributions and mathematical tables and formulas. With an emphasis on statistical reasoning, our user-friendly textbook continues to be written in an informal, conversational style. We include a few carefully chosen mathematical cartoons and photographs with educational captions with hopes that they will make using the book more enjoyable and less frightening to students.

## Homework and Test Problems/Instructor's Manual

There are numerous problems at the end of each chapter. Some give practice in problem-solving technique; others require the use of critical judgment. Answers to the odd-numbered problems are in an appendix at the end of the text. Answers to the even-numbered problems are in the Instructor's Manual so that instructors may assign the problems as homework or in-class exercises. The Instructor's Manual also contains over 700 additional questions. They include several hundred multiple-choice questions for quizzes and several data sets in each chapter appropriate for examinations.

The Instructor's Manual is available to instructors at www.wiley.com/college/king.

## Acknowledgments

Many reviewers, mentors, colleagues, friends, and family members helped develop the text. To them we extend our lasting gratitude for debts that can never be fully repaid. In particular, we thank Ed Minium, who was sole author on this book when it was first published in 1970. His vision and understanding of statistics have guided us ever since. Bruce King thanks his wife, Gail, for her support during the seemingly never-ending time they juggled the demands of authorship, teaching, and research. Patrick Rosopa would like to thank his wife, Melodee, for her patience and support. He would also like to acknowledge his daughters, Elenah and Cecilia, whose love serves as a constant source of strength. He would also like to acknowledge his Filipino immigrant parents for their love, for their hard work, and for stressing the importance of education. We are especially grateful for the leadership of Veronica Visentin and Judy Howarth, whose diligence and patience enabled us to complete the project.

In addition, we want to express our gratitude to the following reviewers for suggestions made during various stages of the manuscript. We may not have made every change the reviewers desired, but each suggestion caused us to stop and carefully reassess our goals and the needs of the students and the instructors. The reviewers are as follows:

Robert W. Allan, Lafayette College

Jeffrey Andre, James Madison University

Razia Azen, University of Illinois at Urbana–Champaign

Frank M. Bagrash, California State University–Fullerton

Dr. D. K. Beale, Cerrito College

Dr. David M. Boynton, Western Connecticut State University

Shirley Brown, SUNY Albany

Jerry Cohen, University of Rhode Island

Jane Conner, SUNY Binghamton

Stephen S. Cooper

Donna Cruse

Ernest C. Davenport, University of Minnesota–Twin Cities

Jessica Dennis, California State University, Los Angeles

Janis W. Driscoll, University of Colorado at Denver

Bruce Dudek, SUNY Albany

Rick M. Gardner, University of Colorado at Denver

Sonia V. Gonsalves, Richard Stockton College of NJ

Sue A. Joslyn, University of Northern Iowa

Shelia Kennison, Oklahoma State University

Dr. Melanie Kercher, Sam Houston State University

Donald L. Meyer, University of Pittsburgh

Frederick L. Oswald, Michigan State University

Robert R. Provine, University of Maryland–Baltimore County

Robert M. Pruzek, SUNY Albany

William P. Reich, Loyola University of Chicago

Alan Reifman, Texas Tech University

William M. Rogers, Grand Valley State University

Margaret G. Ruddy, The College of New Jersey

Ron Serlin, University of Wisconsin–Madison

Murray Singer, University of Manitoba

Gregory T. Smith, University of Kentucky

Alan Sockloff, Temple University

John E. Stecklein, University of Minnesota

Dr. Chehalis Strapp, Western Oregon University

Lawrence J. Stricker, The New School for Social Research

Ellen P. Susman, Metropolitan State College

Penelope G. Vinden, Clark University

Julia Wallace, University of Northern Iowa

William Yost, Loyola University of Chicago

Dr. Lynnette Zellezny, California State University at Fresno

## About the Book and Authors

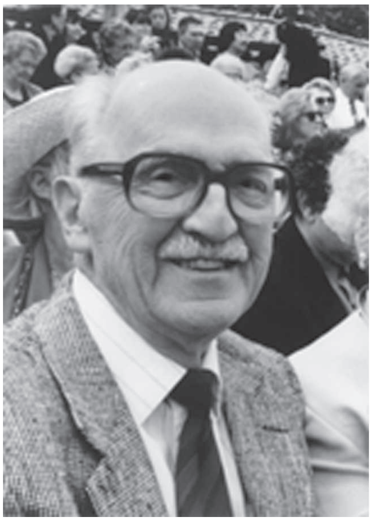
This book was first published in 1970 with the title *Statistical Reasoning in Psychology and Education*. The sole author was Edward Minium. Bruce King assumed the responsibility of revising the book for the third edition, published in 1993. The book's title changed to *Statistical Reasoning in the Behavioral Sciences* for the fifth edition, published in 2008. Patrick Rosopa co-authored the sixth edition and will assume the primary responsibility for revisions in future editions.

Bruce M. King 1946–

Bruce King earned his B.A. in psychology at UCLA in 1969 and a Ph.D. in biopsychology at the University of Chicago in 1978. He was a faculty member at the University of New Orleans from 1978 to 2007 and is presently at Clemson University. He has been honored as a Fellow in the Association for Psychological Science, American Psychological Association, and International Behavioral Neuroscience Society.

Patrick J. Rosopa 1972–

Patrick Rosopa earned his B.S. (2000) in psychology at Tulane University and his M.S. (2003) and Ph.D. (2006) in industrial/organizational psychology at the University of Central Florida. He is an Associate Professor at Clemson University. He has published in various peer-reviewed journals including *Psychological Methods, Organizational Research Methods, Journal of Modern Applied Statistical Methods, Journal of Managerial Psychology*, and *Human Resource Management Review*. Currently, he serves on the editorial board of *Organizational Research Methods, Journal of Managerial Psychology*, and *Research in Human Resource Management*.



Edward W. Minium 1917–2004

Ed Minium earned his B.A. at Stanford University in 1939. After serving with the United States Air Force during World War II, he returned to earn his Ph.D. in psychology at the University of California at Berkeley in 1951. He was a faculty member at San José State University from 1948 to 1982 and served as chair of the Department of Psychology from 1961 to 1966. He was the only author for the first two editions of this textbook, which was first published in 1970.

# Contents

## 14 Interpreting the Results of Hypothesis Testing: Effect Size, Type I and Type II Errors, and Power 205



## 15 Testing Hypotheses about the Difference between Two Independent Groups 224

## 16 Testing for a Difference between Two Dependent (Correlated) Groups 249



## 17 Inference about Correlation Coefficients 264



## 18 An Alternative to Hypothesis Testing: Confidence Intervals 271

## 19 Testing for Differences among Three or More Groups: One-Way Analysis of Variance (and Some Alternatives) 289



## 20 Factorial Analysis of Variance: The Two-Factor Design 326

# Introduction

# 1

When you have finished studying this chapter, you should be able to:

- Define statistics as a specialization within the field of mathematics;
- Explain the difference between descriptive and inferential statistics;
- Understand that the term *statistic* has a second meaning—a descriptive index of a sample—and know the difference between a statistic and a parameter;
- Explain the role of applied statistics;
- Understand the difference between a research question and a statistical question and the difference between a statistical conclusion and a research conclusion;
- Understand that statistical techniques are tools that are only as good as the individual who uses them, but that sometimes even good researchers make incorrect statistical conclusions;
- Understand the difference between variables and constants and the difference between a discrete variable and a continuous variable;
- Recognize measurements as being on one of four scales—nominal, ordinal, interval, or ratio—and understand that not all numbers can be treated alike; and
- Appreciate that to learn statistics you may have to read and study differently than for other subjects.

Is statistics important to people's lives? Consider the following two research studies.

**Example 1**

In early 1987, the U.S. Food and Drug Administration (FDA) was faced with an unprecedented situation. Thousands of people were dying of acquired immunodeficiency syndrome (AIDS), a relatively new sexually transmitted infection caused by the human immunodeficiency virus. The number of new cases of AIDS was increasing so rapidly that the Public Health Service was predicting that more than 50,000 Americans would soon be dying of the infection every year. Not only was there no known cure, but there was not even a drug available to slow the progression of the infection. However, early clinical trials of an experimental antiviral drug known then as azidothymidine (AZT) were promising. Only 1 of 145 AIDS patients on AZT had died, compared to 19 of 137 patients in a control group given a placebo (a substance that is medically without effect).

The process of approval of a new drug is typically very slow (8.8 years, on average), and several more years of clinical trials were scheduled. There were medical questions remaining to be answered. What was the optimal dose? For how long would the drug continue to thwart the virus? Would it prove to have toxic side effects? There were ethical questions as well. Was it morally right to withhold from other patients a drug that seemingly prolonged lives? There was also an important statistical question, one that had to be answered before the medical and ethical questions could be addressed. Among AIDS patients using AZT, was the fewer number of deaths the result of the drug, or was it due just to chance? Even if the FDA had given both groups of AIDS patients the placebo or both AZT, it would be unlikely that the same number of deaths would have occurred in both groups. One would expect a difference due to chance. Were the observed differences in deaths great enough to warrant early approval of the drug? Statistical tests showed that the difference between the two groups was so great that *the probability of it having occurred by chance was less than 1 in 1,000* (Fischl et al., 1987). Armed with these statistics, the FDA gave final approval for the use of AZT in March 1987 after only 21 months of testing. The FDA's decision to give early approval for AZT was an ethical one designed to prolong lives, but it was made only after statistical analysis.

**Example 2**

For women, the ovaries are a source of two important hormones: estrogen and progesterone. As women approach age 50, their supply of eggs within the ovaries nears zero, they have their last menstrual period (called menstruation) and the ovaries atrophy. The loss of estrogen and progesterone was believed to cause many serious problems (including hot flashes and weaker bone density), and thus for many years the standard medical treatment for postmenopausal women was hormone replacement therapy.

In 1997, a large study was begun to compare the long-term effects of a popular hormone replacement drug (Prempro) versus a placebo. It was expected that Prempro would have very beneficial effects in postmenopausal women. The study was originally designed to last 8 years, but it was ended after 5 years. The reason was that the women given Prempro were experiencing a greater number of heart attacks, strokes, blood clots, and breast cancers (Manson et al., 2003). Statistical analysis had shown that the greater number of bad outcomes experienced in the group given Prempro compared to the group given the placebo *had a low probability of occurring by chance* (i.e., it was more likely due to the hormone replacement therapy). The results of this study have been a major influence on how postmenopausal women receive medical treatment today.

Statistics is used not only in making important decisions, as in the cases of AZT and hormone replacement therapy, but also in our everyday lives. When your professor organizes test scores into a grade distribution and calculates the class average, he or she is using statistics. If you have ever kept track of the number of gold, silver, and bronze medals won by athletes in Olympic games competition, you have used statistics. Without statistics, data collected in our everyday observations or in carefully controlled experiments would have very little meaning.

What do we mean when we use the word *statistics*? In ordinary language, most people use the term *statistics* to refer to any set of facts that involve numbers (e.g., "birth statistics," "crime statistics," "unemployment statistics"). When used like this, *statistics* is a plural word (e.g., "The crime statistics are going to be reported on Wednesday"). However, in this section we refer to *statistics* as a science, a specialization within the field of mathematics. **Statistics** is the science of classifying, organizing, and analyzing data. In this sense, the word is singular (e.g., "Statistics is a science"). We introduce yet another meaning of the term *statistics* in Section 1.2.

**statistics** the science of classifying, organizing, and analyzing data (for another meaning, see Section 1.2)

The goal of this book is to teach you the techniques of statistics used by researchers to analyze data. Some of these techniques will be very simple, and others will be more complex. In learning statistics, it is useful to divide the subject into two parts: descriptive statistics and inferential statistics.

## 1.1 Descriptive Statistics

In a new school, a biology instructor is thinking about what the first group of students will be like. How much biological information will they already have? She does not wish to bore them by underestimating their knowledge, but she also does not want to lose them by assuming too much. Should she begin with a review of fundamentals? How extensive? In what areas? It would also be helpful to know whether the students vary widely in the extent of their biological knowledge. If they do, the instructional method may need adjustment.

Let us suppose that the biology instructor administers a nationally developed test of biological knowledge to her students at the first class meeting. Let us also assume that she is able to learn how students in similar educational circumstances perform on this test. She discovers, for example, that Bob Beaver's score in zoology is better than that of 90% of students in the same grade and that only 15% of students get lower scores than Steve Smith's.

Because the names in her gradebook are in alphabetical order, the scores are not in numerical order, so she first organizes the scores in descending order and then finds the class average for each subtest and for the complete test. Then she compares the class averages with the performance data for other similar students. She finds that as a group her students are approximately at the expected level in botany and that their performance is superior in zoology. This knowledge will help her go about teaching.

In reorganizing and comparing her students' scores, our instructor is making use of techniques that are part of the body of descriptive statistics. The purpose of **descriptive statistics** is to organize and to summarize observations so that they are easier to comprehend.

**descriptive statistics** its purpose is to organize and summarize observations

**inferential statistics** its purpose is to draw a conclusion about conditions that exist in a population from study of a sample

## 1.2 Inferential Statistics

What is the attitude of the voting public toward capital punishment? A second branch of statistical practice, known as inferential statistics, provides the basis for answering questions of this kind. Pollsters would find it impossible to ask all registered voters a question like the one about capital punishment, so they ask only a small portion of voters and then use the answers to estimate the attitudes of all voters. The purpose of **inferential statistics** is to draw a conclusion (an inference) about conditions that exist in a **population** (the complete set of observations) by studying a **sample** (a subset) drawn from the population. Because of chance factors associated with drawing a sample, the outcome, like any estimate, is subject to error. However, if the sample of voters selected for study has been chosen according to statistical principles, it is possible to know what margin of error is involved.

**population** the complete set of observations about which an investigator wishes to draw conclusions

**sample** a subset of a population

We previously defined statistics as a science. However, the term statistic has another meaning that we will use throughout the book. A **statistic** is a descriptive index of a sample. The same index, if descriptive of a population, is called a **parameter**. Thus, the average of a sample of scores is a statistic; the average of a population of scores is a parameter. *In inferential statistics, we will be interested in estimating population parameters from sample statistics*.

**statistic** a descriptive index of a sample (for another meaning, see page 2.)

**parameter** a descriptive index of a population

Here is another example of how inferential statistics is used to evaluate the outcome of an experiment. Is it possible that a certain drug has an effect on the speed of learning? Let us suppose that an investigator decides on the kind of subjects she wishes to study. As in the first example, it is impossible to administer the drug to everyone in the population, so the investigator selects at random two samples of 25 subjects each. **Random sampling** is a procedure that ensures that all samples of a particular size have an equal chance of being selected and thus eliminates any bias

**random sampling** a procedure that ensures that all samples of a particular size have an equal chance of being selected when we draw a sample. The investigator then administers the drug to one of the groups and a placebo to the other group. She gives both groups a learning task and treats both alike in all ways. She finds that the average learning scores of the two groups differ by 5 points.

We would expect some difference between the groups even if both received the drug because of chance factors involved in the random selection of the groups. The question that the experimenter must ask herself is whether the observed difference is within the limits of expected variation. If certain preconditions have been met, statistical theory can provide the basis for an answer. If the experimenter finds that the obtained difference of 5 points is larger than can be accounted for by chance variation, she will infer that other factors must be at work. If examination of her experimental procedures reveals no reasonable cause for the difference other than the deliberate difference in the experimental treatment, she may make an inference about the effects of the drug in the population. This is the same type of reasoning that the FDA investigators used to conclude that AZT prolonged the lives of patient with AIDS. Their conclusion was not about just those patients in the study. It was about the effects of AZT if it were given to *all* persons with AIDS.

## 1.3 Our Concern: Applied Statistics

Statistics originated from the work of mathematicians and others who were well grounded in mathematics. However, the mathematical development of statistics is not what this book is about. Instead, we will be concerned with applied statistics. Our outlook will be that of investigators who need to know statistics to appreciate reports of findings in their professional fields or who must apply statistics in their own work. We may think of ourselves as statistical technicians, therefore. This includes biologists, educators, psychologists, engineers, sociologists, medical researchers, and business executives. All these professionals and many more regularly find that statistical procedures can be helpful in their work.

*Applied statistics is a tool and neither a beginning nor an end in itself.* For any subject area, scientific study begins with an investigator posing a question based on the results of previous studies and theoretical considerations. This is the **research question**, a question of fact concerning the subject matter under investigation. Suppose, for example, that we are interested in whether reaction time depends on the intensity of a stimulus.

**research question** a question of fact about the subject matter under investigation

We must next decide on the specifics necessary to explore the question. We may decide to assign 50 available subjects at random to two groups of 25, one of which will receive a strong stimulus and the other a weaker one. Other designs are possible, but the experimental design will depend, for the most part, on the research question. For example, if we believe that reaction time decreases as stimulus intensity increases, we could test reactions to more than two intensities of the stimulus and assign fewer subjects to a larger number of groups.

**independent variable** the variable that is systematically manipulated by the investigator

**dependent variable** the variable that is measured

**confounding variables** factors (other than the independent variable) that might affect the dependent variable

The variable that is systematically manipulated by the investigator is called the **independent variable** (e.g., stimulus intensity). The variable that is measured is called the **dependent variable** (e.g., reaction time). It is important that all other factors that might affect the dependent variable (**confounding variables**) be the same for all groups. For example, it would be important to see that the subjects have been tested at approximately the same time of day to avoid having one group fresh and alert while the other is tired and inattentive. Having designed our study, we are finally ready to make observations. We measure the reaction time of each subject and record it.

**statistical question** a question about a numerical aspect of the observations

**statistical conclusion** a conclusion about a numerical property of the data

After recording our measurements, we calculate the average reaction time for each group, see that they differ, and raise a statistical question: Is the average reaction time under the two conditions so different that chance variation alone cannot account for it? A **statistical question** is a question about a numerical aspect of the observations. After applying a statistical procedure, we arrive at a statistical conclusion. A **statistical conclusion** is a conclusion about a numerical property of the data. For example, as a result of applying the appropriate statistical procedures in our hypothetical experiment, we might be able to conclude that the average reaction times under the two conditions are so different that it is not reasonable to believe that chance alone could account for it.

Now we return to the research question. If we had carefully arranged the conditions of the study, it may be possible to conclude that intensity of the stimulus does make a difference, at least under the conditions that held in this experiment. This is our final step, and it is a research conclusion. A **research conclusion** is a conclusion about the subject matter. In our initial example about AZT, the statistical conclusion was that the number of deaths in the two groups differed by beyond what would have been expected by chance. The research conclusion was that the difference was due to AZT. Although the research conclusion derives partly from the statistical conclusion, we see that other factors must be considered. The investigator, therefore, must weigh both the statistical conclusion and the adequacy of the research design in arriving at the research conclusion.

**research conclusion** a conclusion about the subject matter

It is important to understand that **statistical procedures** *are only a middle step in the investigation of a research question.* If an investigator has not carefully controlled for all confounding variables, it is possible to make a correct statistical conclusion and at the same time make an incorrect research conclusion. Such was the case in a well-known study by Joseph Brady and some colleagues at Johns Hopkins University (Brady, Porter, Conrad, & Mason, 1958). They investigated the effects of stress on the development of ulcers in four pairs of monkeys. Each pair was locked into adjacent chairs during testing. One monkey in each pair, designated the "executive" monkey, could avoid a strong electric shock by pressing a lever that postponed the shock for 20 seconds. Whenever the executive monkey did receive a shock, the other monkey in the pair, called the "yoked" control monkey, also received a shock (which it could not avoid). The investigators assigned monkeys to the executive group based on their high rates of lever pressing in previous studies.

**statistical procedures** only a middle step in the investigation of a research question

The executive monkeys developed ulcers, but the yoked monkeys did not, even though they received identical shocks. The difference was so great that it was unlikely that the results occurred by chance (the statistical conclusion). Because they believed that they had controlled for all other factors, Brady and his colleagues concluded that it was the continuous stress of having to decide how often to press the lever to avoid shock that caused the ulcers in the executive monkeys (the research conclusion).

Later studies were unable to replicate the results. Moreover, it was found that animals selected for high rates of responding were more susceptible to developing ulcers than were other animals (Weiss, 1971). When monkeys that did not differ in activity level were tested together, the executive monkeys developed less severe ulcers (Foltz & Millett, 1964; Natelson, 1976). We now know that in general, control over an aversive event reduces the stress the event causes (Taylor, 1991). The original investigators had unknowingly biased the outcome of their experiment by not randomly assigning the monkeys to the two conditions and thus reached an incorrect research conclusion.

## 1.4 Variables and Constants

When we conduct a study, we must often consider several variables and constants. A **variable** is a characteristic that may take on different values. Typical examples are intelligence test scores, number of errors on a spelling test, height, marital status, and gender. The concept of a variable does not imply that each observation must differ from all the others. All that is necessary is the possibility of difference. Suppose, for example, that a school nurse is interested in the height of seventh grade boys in Lincoln Junior High School. If she selects a sample of three for study and finds that they are all the same height, it is still proper to refer to height as a variable because the possibility of getting students of different height existed.

**variable** a characteristic that may take on different values

On the other hand, her decision to study height only among the seventh grade students means that grade level is a constant rather than a variable in this inquiry. When, in terms of a particular study, it is not possible for a characteristic to have other than a single value, that characteristic is a **constant**. Constants limit the applicability of the results of a study. In the school nurse's situation, for example, the sex of the students (male), the school (Lincoln), and the grade level (seventh) are all constants. Even if the nurse had taken a larger sample, the conclusions that she would draw could apply for sure only to seventh grade boys in that school.

**constant** a characteristic that can have only one value (other values are not possible)

Variables may be either discrete or continuous. A **discrete variable** can take on only certain values. If we count the people at a meeting, we might learn that 73 persons attended, or 74, but no value between these two figures is possible. The values are *exact numbers.* Other examples include the number of multiple-choice questions answered on an exam or the number of lever presses performed by a rat in a learning experiment.

**discrete variable** a variable that can take on only certain values

A **continuous variable** can take on any value (within whatever limits its values may range). In measuring length, for example, it is possible for an object to be 3 ft. 1 in. or 3 ft. 2 in. long, or any conceivable length in between. A continuous variable has no gaps in its scale. Other examples include weight, age, and temperature.

**continuous variable** a variable that can take on any value (within the limits that its values may range)

*Even though a variable is continuous in theory, the process of measurement always reduces it to a discrete one*. For example, if we measure weight to the nearest tenth of a pound, a weight of 18.3 pounds means that the object is closer to 18.3 pounds than it is to 18.2 pounds or 18.4 pounds. The value of 18.3 pounds is an *approximate number*. The actual weight is somewhere between 18.25 and 18.35 pounds. In this example, our recorded measurements form a discrete scale in steps of one-tenth of a pound. When measuring a continuous variable, our level of accuracy is limited by our recording equipment, and it is up to the investigator to determine the degree of precision appropriate to the problem at hand. Your height may be 67 and 7/16 in., but who cares? For most of us, measurement to the nearest half-inch is fine.

Typical statistical computations involve both exact and approximate numbers—for example, dividing the sum of the scores (usually the sum of approximate numbers and thus itself approximate) by the number of cases (an exact number) to produce an average. *In computations involving both exact and approximate numbers, the accuracy of the answer is limited by the accuracy of the approximate numbers involved*. For example, if we measure weight to the nearest tenth of a pound and find that the total weight of three objects is 67.0 pounds, it would be misleading to report the average weight (67.0/3) as either 22.333333 pounds (the output of the common eight-digit display calculator) or 22 pounds. Because the approximate number is reported to the nearest tenth of a pound, the reported average weight should be 22.3 pounds.

When working with a series of statistical computations involving approximate numbers, it is a good idea to keep a little more accuracy than we think is the minimum to which we are entitled (perhaps one extra decimal place) because in a sequence of computations it is possible to compound inaccuracy. Once the computation is completed, we should round back to the least accurate of the numbers used. If the last digit is greater than 5, round up; if it is less than 5, round down. *When rounding a number that ends exactly with the numeral 5, you should always round to the nearest even number*. This avoids introducing a bias in your calculations. (Sometimes you will round up and other times down rather than always in the same direction.) For example, if rounding to the nearest tenth, 1.35 becomes 1.4, as does 1.45.

## 1.5 Scales of Measurement

**Measurement** is the process of assigning numerals to observations. This is not done arbitrarily, but in a way so that the numbers are meaningful. How do we do this? Variables do not all have the same numerical properties. In measuring weight, we are accustomed to the idea that 40 pounds is twice as much as 20 pounds and that the difference between 10 and 20 pounds is the same as

**measurement** the process of assigning numerals to observations that between 60 and 70 pounds. Similarly, if a baseball team has won 40 games and another only 20, we can say that the first has won twice as many games as the second team. However, if we put baseball teams in order of their standing in the league, we should not think that the number one team is twice as good as the number two team, nor would we expect that the difference in abilities between the first and second team is necessarily the same as that between the second and third team. It is apparent that numbers have a different significance in these two situations. To help distinguish different kinds of situations, the psychologist S. S. Stevens (1946) identified four different scales of measurement. You will find that numbers are treated differently depending on their scale of measurement.

## Nominal Scale

Some variables are qualitative in their nature rather than quantitative. For example, eye color, types of cheese, and party of political affiliation are examples of *qualitative variables*. The several categories of such a variable form a **nominal scale** (the Latin root of *nominal* means "name"). The only requirements of a nominal scale are that the categories be *mutually exclusive* (the observations cannot fall into more than one category) and *exhaustive* (there must be enough categories for all the observations). Brown and blue eyes are mutually exclusive categories for eye color; but Ford, General Motors, Toyota, Chevrolet, and Honda are not mutually exclusive categories for types of cars (Chevrolets are made by General Motors). With a nominal scale, there is no question about one category having more or less of any particular quality; all categories are simply different. We may use numbers to identify the categories of a given scale. For example, we might identify three cheeses as Cheese No. 1, Cheese No. 2, and Cheese No. 3. Numbers used in this way are simply a substitute for names and serve only for purposes of identification. Numbers on football jerseys similarly indicate only the position of the player (e.g., offensive line, offensive backfield). If one football player wears the number 10 on his back while another wears the number 20, there is no implication that the second player is "more than" the other in some dimension, let alone that he has twice as much of something (Stine, 1989).

**nominal scale** mutually exclusive and exhaustive categories differing in some qualitative aspect

## Ordinal Scale

At the next level of complexity is the **ordinal scale** (the Latin root means "order"). In this type of measurement, *the categories must still be mutually exclusive and exhaustive, but they also indicate the order of magnitude of some variable.* With a nominal scale, the outcome of classification is a set of unordered categories. With the ordinal scale, it is a set of ranks. A classic example is military rank. Sergeant is greater than corporal, which is in turn greater than private. Another example is academic standing—freshman, sophomore, junior, and senior. In yet another example, a supervisor may estimate the competence of seven workers by arranging them in order of merit. The only relation expressed in a series of numbers used in this way is that of "greater than." We may use numbers for the ranks but it is not necessary. Among persons ranked 1, 2, and 3, the first person has a greater degree of merit than the person ranked second, and the second person has greater merit than the third. However, nothing is implied about the magnitude of difference in merit between adjacent steps on the scale. Furthermore, nothing is implied about the absolute level of merit; all seven workers could be excellent, or they could be quite ordinary.

**ordinal scale** has the properties of a nominal scale, but in addition the observations may be ranked in order of magnitude (with nothing implied about the difference between adjacent steps on the scale)

## Interval Scale

The next major level of complexity is the **interval scale**. This scale has all the properties of the ordinal scale, but with the further refinement that a *given interval (distance) between scores has*

**interval scale** has all the properties of an ordinal scale, and a given distance between measures has the same meaning anywhere on the scale


**FIGURE 1.1** Three temperatures represented on the Celsius and Kelvin scales.

*the same meaning anywhere on the scale*. Thus, it could be better called an *equal-interval scale*. Examples of this type of scale are degrees of temperature on the Fahrenheit or Celsius scales. A 10° rise in a reading on the Celsius scale represents the same change in heat when going from 0° to 10° as when going from 20° to 30°. The limitation of an interval scale is that it is not possible to speak meaningfully about a ratio between two measurements. We illustrate this point in Figure 1.1. The top of this illustration shows three temperatures in degrees Celsius: 0°, 50°, and 100°. It is tempting to think of 100°C as twice as hot as 50°. However, the value of zero on this scale is simply an arbitrary reference point (the freezing point of water) and does not imply the absence of heat. Therefore, it is not meaningful to assert that a temperature of 100° Celsius is twice as hot as one of 50° or that a rise from 90° to 99° Celsius is a 10% increase.

## Ratio Scale

**ratio scale** has all the properties of an interval scale plus an absolute zero point

A **ratio scale** possesses all the properties of an interval scale and in addition *has an absolute zero point*. The bottom of Figure 1.1 shows the same three temperatures in degrees Kelvin. This scale uses the same unit for its intervals; for example, 50° of change in temperature is the same on both scales. However, the Kelvin scale has an absolute zero, the point at which a substance would have no molecular motion and, therefore, no heat. Thus, 100° is twice as hot as 50° on the Kelvin scale (100° is only 1.15 times as hot as 50° on the Celsius scale). Other examples of ratio scale measurements are length, weight, and measures of elapsed time. Not only is the difference between 40 and 41 in. the same as the difference between 80 and 81 in., but it is also true that 80 in. is twice as long as 40 in.

Table 1.1 summarizes the characteristics of the four scales of measurement we have just discussed.

# 1.6 Scales of Measurement and Problems of Statistical Treatment

An understanding of the scales of measurement provides a framework for appreciating some problems in the interpretation of data. The first thing we need to realize is that in the behavioral sciences, there are many measuring instruments that lack equal intervals and an absolute zero point. Consider, for example, a spelling test. Easier items might be words such as *garden*, *baseball*, and *rowboat*. But we might also find *perceive, complacency*, and perhaps even some stumpers like *gauge*, *accede*, and *abscissa.* A score of zero on this test means that the person could not spell the simplest word on the list, but what if simpler words had been on the test, such as *cat*, *run*, and *bat*? Our spelling test, then, does not have an absolute zero point because zero on

**Table 1.1 Scales of Measurement and Their Characteristics**

| SCALE | PROPERTIES | EXAMPLES |
|---|---|---|
| Nominal | Mutually exclusive and exhaustive categories differing in some qualitative aspect. | Sex, ethnic group, religion, eye color, academic major |
| Ordinal | Scale has the property of a nominal scale (mutually exclusive categories) and in addition has observations ranked in order of magnitude. Ranks, which may be numerical, express a "greater than" relationship, but with no implication about how much greater. | Military rank, academic standing, workers sorted according to merit |
| Interval | Scale has all the properties of an ordinal scale, and in addition, numerical values indicate order of merit and meaningfully reflect relative distances between points along the scale. A given interval between measures has the same meaning at any point in the scale. | Temperature in degrees Celsius or Fahrenheit |
| Ratio | Scale has all the properties of an interval scale, and in addition has an absolute zero point. Ratio between measures becomes meaningful. | Length, weight, elapsed time, temperature in degrees Kelvin |

the spelling test does not indicate a total absence of spelling ability. The same is true of midterm tests, IQ tests, the SAT, and almost all other tests of mental performance.

What about equal intervals? To have equal intervals on our spelling test, we should be able to state quantitatively just how much more spelling ability is needed to spell *garden* than to spell *cat* and how much more is needed to spell *gauge* than to spell *rowboat*. So far, we have found no objective way to determine when equal numerical intervals on a mental test represent equal increments in performance. Realization of this situation once caused great concern in statistical circles. Some people argued that calculating certain statistical variables (such as averages) on tests of mental abilities could be seriously misleading. Fortunately, the weight of the evidence suggests that in most situations, making statistical conclusions is not seriously hampered by uncertainty about the scale of measurement (Norman, 2010). We will consider this matter further in the Point of Controversy in Chapter 4.

There are, however, several areas where we need to be aware of scale problems to avoid taking tempting but incorrect positions. For example, we should not say that a person with an IQ of 150 is twice as bright as one with an IQ of 75, or that the difference between 15 and 25 points on a spelling test necessarily represents the same increment in spelling ability as the difference between a score of 30 and 40 points on the same test. In psychological measurement, this problem may be particularly critical when a test does not have enough "top" or "bottom" to differentiate adequately among the group measured. For example, imagine a test of ability that has a maximum possible score of 50 points and that is too easy for the group measured. For two persons who score 50 points, the score for one may indicate the maximum level of achievement, but the second person may be capable of a much higher level of performance. The measuring instrument is simply incapable of showing this difference because it does not include items of greater difficulty.

## 1.7 Do Statistics Lie?

> **There are three kinds of lies: lies, damned lies, and statistics.**
> Benjamin Disraeli, 19th-century British statesman

Do statistics really lie? In the experiment by Brady and his colleagues, the investigators reached an incorrect research conclusion, but there really was an observed difference between the executive and yoked monkeys in their susceptibility to developing ulcers. Is it possible to