FIFTH EDITION

STATISTICS

An Introduction

Roger E. Kirk
Baylor University

THOMSON
WADSWORTH

Australia • Canada • Mexico • Singapore • Spain
United Kingdom • United States

THOMSON
WADSWORTH

Statistics
Roger E. Kirk

Publisher: Michele Sordi
Assistant Editor: Gina Kessler
Editorial Assistant: Christina Ganim
Technology Project Manager: Lauren Keyes
Marketing Manager: Karin Sandberg
Marketing Assistant: Natasha Coats
Senior Marketing Communications Manager: Linda Yip
Content Project Manager: Karol Jurado
Creative Director: Rob Hugel
Senior Art Director: Vernon Boes
Print Buyer: Doreen Suruki

Permissions Editor: Sarah D'Stair
Production Service: Pre-Press Company, Inc.
Text Designer: John Edeen
Copy Editor: Karen Carriere
Illustrator: Pre-Press Company, Inc.
Cover Designer: Brenda Duke Design
Cover Image: Hoberman Collection (from Photonica)
Cover Printer: RR Donnelley Crawfordsville
Compositor: Pre-Press Company, Inc.
Printer: RR Donnelley Crawfordsville

Printed in the United States of America
1 2 3 4 5 6 7 11 10 09 08 07

Library of Congress Control Number: 2006933545

ISBN-13: 978-0-534-56478-0
ISBN-10: 0-534-56478-X

Thomson Higher Education
10 Davis Drive
Belmont, CA 94002–3098
USA

About the Author

Roger E. Kirk received his Ph.D. in experimental psychology from the Ohio State University and did postdoctoral study in mathematical psychology at the University of Michigan. He is a Distinguished Professor of Psychology and Statistics at Baylor University. He founded and, for 25 years, directed Baylor's Behavioral Statistics Ph.D. program and the Institute of Statistics, now the Department of Statistical Science. He has published extensively in the areas of statistics, psychoacoustics, and human engineering and is the author of five statistics books. His first book, *Experimental Design: Procedures for the Behavioral Sciences*, has been identified by the Institute for Scientific Information as one of the most frequently cited books in its field. Dr. Kirk is a fellow of the American Psychological Association (Divisions 1, 2, 5, and 13) and the American Psychological Society. He is a past president of the Society for Applied Multivariate Research, Division 5 of the American Psychological Association, and the Southwestern Psychological Association. In recognition of his teaching effectiveness, he was named the Outstanding Tenured Teacher in the College of Arts and Sciences and designated a Master Teacher, Baylor University's highest teaching honor. He is the 2005 recipient of the Jacob Cohen Award for Distinguished Contributions to Teaching and Mentoring from Division 5 of the American Psychological Association.

1.1 INTRODUCTION

Looking Ahead: What Is This Chapter About?

When a student came to me recently for help with statistics, I posed the question, "What is the chapter about?" The student's answer, "About 36 pages," was not what I had hoped to hear. To give you a heads up, I provide a brief overview at the beginning of each chapter.

This chapter begins with a discussion of what statistics is and why you should study it. I then share tips for studying statistics and define some basic concepts: population, sample, and random sample. You will learn that there are two broad categories of statistics: descriptive statistics and inferential statistics. The chapter continues with a discussion of the way mathematicians classify variables and the rules psychologists and others use to assign numbers to characteristics of people. For history buffs, I end the chapter with a brief description of the origins of statistics.

After reading the chapter, you should know the following:

- What statistics is
- Why you should study it (although you might prefer almost any other form of torture)
- How to study statistics
- The meaning of basic concepts such as population, sample, and random sample
- The two broad categories of statistics
- The way mathematicians classify variables and the way psychologists measure characteristics
- The origins of statistics

Some Misconceptions

It is widely believed that statistics can be used to prove anything—which implies, of course, that it can prove nothing. Furthermore, the word *statistics* conjures up visions of numbers piled upon numbers, uninterpretable charts, and computers cranking out gloomy predictions. To the ordinary person, besieged from all sides by advertising claims, statistics is hocus-pocus with numbers. It was Benjamin Disraeli who said, "There are three kinds of lies—lies, damned lies, and statistics."[1] In primitive cultures, exaggeration was common. One writer, with tongue in cheek, reasoned that because primitive people did not have a science of statistics, they were forced to rely on exaggeration, which is a less effective form of deception. Another writer remarked, "If all the statisticians in the world were laid end to end—it would be a good thing." Whatever its public image, statistics endures as a required course, and my students continue to refer to it, affectionately no doubt, as Sadistics 2402.

[1] Three books indicate that Disraeli's view of statistics is still with us: *How to Tell the Liars from the Statisticians* by Hooke and Liles, *Misused Statistics: Straight Talk for Twisted Numbers* by Jaffe and Spirer, and *Statistical Deception at Work* by Mauro.

What Is Statistics?

In spite of frequent misuse, statistics can be a powerful tool for making decisions in the face of uncertainty. The word *statistics* comes from the Latin *status*, which is also the root for our modern term *state* or political unit. Statistics was a necessary tool of the state, because to levy a tax or to wage war a ruler had to know the number of subjects in the state and the amount of their wealth. Gradually the meaning of the term expanded to include any type of data.

> Today the word **statistics** has four distinct meanings. Depending on the context, it can mean (1) data; (2) functions of data, such as the mean and range; (3) techniques for collecting, analyzing, and interpreting data for subsequent decision making; and (4) the science of creating and applying such techniques.

Why Study Statistics?

A knowledge of statistics yields more than the obvious benefits. For example, it generates new ways of thinking about questions and effective tools for answering them. It takes only a cursory examination of the professional literature in your field to see the inroads made by statistical techniques and ways of thinking. Statistics helps researchers make sense of data and is an indispensable research tool, but its usefulness is not limited to research. In many fields, it is virtually impossible to read research articles and keep up with new developments without an understanding of elementary statistics. Also, statistics is an interesting subject—some people even find it fascinating.

In all likelihood, you are reading this book because it was assigned in your required statistics course. You have been told that the study of statistics is necessary, and there is a strong implication that it will be good for you. At this point you may be skeptical. Just what can you expect to learn by studying statistics? A quick scan of this book will give you an idea. You will acquire a new vocabulary, because in many ways learning statistics is like learning a foreign language, and you will learn to manipulate numbers according to symbolic instructions. But more important, you will learn when and how to apply statistics to research problems in the behavioral sciences, health sciences, and education. Your study of statistics should enable you to read the literature in your field with greater understanding and make you a more critical consumer of statistical presentations in the mass media. And you should gain a greater appreciation of the probabilistic nature of scientific knowledge. Statistics involves a special way of thinking that can be used not only in research but also in one's daily life. I hope that you will add this way of thinking to your conceptual arsenal.

Kinds of Statisticians

> Users of statistics fall into four categories: (1) those who must be able to read and understand statistical presentations in their field; (2) those who select, apply, and interpret statistical procedures in their work; (3) applied statisticians; and (4) mathematical statisticians.

This book addresses those in the first two categories, including psychologists, educators, speech therapists, biologists, nurses, medical researchers, and physical therapists, to mention only a few. In each case the person's primary interest is in his or her own field, be it counseling or physical therapy; he or she is interested in statistics because it is a useful tool for answering questions in that field. These people are both consumers and users of statistics. Their knowledge of statistics can range from meager to expert.

The applied statistician helps professionals in substantive areas to use statistics effectively. He or she may work for industry or a government agency, engage in a private consulting practice, or teach in a university. Unlike individuals in the first two categories, an applied statistician usually has advanced degrees in statistics.

The mathematical statistician is primarily interested in pure (mathematical) statistics and probability theory rather than in the application of statistics to substantive areas. Most likely this statistician teaches in a university and makes contributions to the theoretical foundations of statistics that may ultimately be used by those with applied interests.

1.2 STUDYING STATISTICS

Develop Effective Study Techniques

Psychologists say that learning is easier when you can integrate new information into an existing knowledge base. Unfortunately, as you begin your study of statistics, your statistical knowledge base is minimal. Building a knowledge base is easier if you use effective study techniques. For example, always survey your reading material by thumbing through the assigned pages and noting topic headings and boldface terms. Try to get a sense of what the material is about. Your survey will provide a general orientation to the material and help you fit facts together as you develop your statistical knowledge base.

Before you begin reading a section in the text, turn the section heading into a question. The question for this section might be, "What are some effective study techniques?" After you have formed your question, look for the answer as you read the section. Research on learning tells us that an active, searching attitude on the part of the reader promotes better learning than does a passive attitude. After reading a section, try to recall the main points of the section by reciting out loud. All of us have had the experience of reading a paragraph and having no idea of what we have just read. Knowing that we will attempt to recall what we are reading develops a mental set to select and retain important facts.

Most forgetting takes place within the first 24 to 48 hours after learning. You can minimize the forgetting process by reviewing your assignment a day or so after reading it. Look at the major headings and boldface terms and see whether you can recite the main points that were covered in the section and define each boldface term. If the contents of some sections are hazy, reread these sections and see whether you can then recall the main points.[2]

[2] These study suggestions are based on the famous SQ3R study method developed by Francis P. Robinson (1946). The letters SQ3R stand for *S*urvey, *Q*uestion, *R*ead, *R*ecite, and *R*eview.

Plan to Read More Slowly

Statistics cannot be read like assignments in history, English, or political science. Ideas and computational procedures in statistics are presented in a highly symbolic form and use a specialized vocabulary that you must learn. Consequently, a 30-page assignment may take three or four times as long to read as a comparable assignment in history. You will understand many sections of this book on a first reading; others will require two or more readings, lots of concentration, and perhaps some time between readings for the ideas to sink in.

Don't Worry If You Weren't an Ace in Math

If you're concerned about the level of mathematics required to understand statistics, stop worrying. Most statistical procedures in this book involve nothing more complicated than addition, subtraction, multiplication, and division. Although this book makes some use of high school algebra, the level is very elementary. For those whose skills are rusty, the essential arithmetic and algebra are reviewed in Appendix A.

Appendix A also contains a diagnostic math test that you can take to assess your math skills and see if you have forgotten anything. I encourage you to check out your skill level by taking the test and grading your performance. I have provided a table of norms based on the scores of my students over the past 10 years.

But don't get too hung up on mathematics. Treat this course less like a math course and more like a course in logic. You should focus on the concepts and the logic underlying statistical procedures. Leave the mathematics and computations to calculators and computers.

Resolve to Review Often

Unless you frequently review this material, it will slip away. Don't skip the *Check Your Understanding* exercises at the end of each section and the end-of-chapter *Review Exercises*. They (1) provide feedback about what you know and what you don't, (2) indicate which concepts and computational procedures are the most important, (3) offer numerous examples of how statistics are used, and (4) give you practice in applying what you are learning. Answers to all of the *Check Your Understanding* exercises are given in Appendix C. The "Looking Back: What Have You Learned?" section at the end of each chapter also is useful for reviewing because it showcases the most important concepts and places the topics in perspective.

The best way to learn statistics is to *do* statistics. By doing the *Check Your Understanding* exercises "by hand" with the aid of a calculator you will gradually learn how to follow the sequence of mathematical operations represented by a formula. Computing a statistic by hand helps to develop an intuitive understanding of the statistic. Once you have an intuitive understanding, it is time to let a computer do the work.

Master Foundation Concepts before Going on to New Material

In statistics, as in mathematics or a foreign language, the material presented first is the foundation for what follows. It is best to master each chapter before you go on to the next. Fight the temptation to cram. Cramming can be effective for some subjects, at least as far as tests are concerned. But in statistics, it inevitably results in a superficial understanding of basic concepts and subsequent learning problems. Periodic reviews require discipline, but they pay off.

Strive for Understanding

This book contains hundreds of formulas. I have not memorized all of them, and neither should you. Some, such as the one for the arithmetic mean, $\overline{X} = \sum X/n$, appear so often that you really can't help learning them; the others aren't worth the effort. I decided a long time ago, when faced with my inability to remember telephone numbers and addresses, that books are better repositories than my head for such things. In all likelihood you will do most of your statistical calculations with computers and calculators. These tools have phenomenal memories for formulas and can spew out statistics at the press of a key.

Instead of memorizing formulas, strive to understand the logic underlying the statistical procedures that you are learning, and think about ways that each new statistic can be applied. In what situations is the statistic useful? How is the statistic interpreted? What assumptions must be fulfilled to interpret the statistic? When you read about an experiment in your field, consider how you would have designed it and how you would have analyzed the data. And check out your ideas by talking about them with your professor and other students. There is no better way to deepen your understanding of a new concept than to explain it to a classmate.

1.3 BASIC CONCEPTS

Population and Sample Defined

Many statistical terms are a legacy from the time when statistics was concerned only with the condition of the state. *Population*, for example, originally meant, and still means, the total number of inhabitants of a state. Its meaning in statistics is broader.

> A **population** is the collection of all people, objects, or events having one or more specified characteristics.

The population is identified when you specify its common characteristics. All the people listed in a telephone directory constitute a population, as does the number of heads and tails obtained in tossing a coin for eternity.

> A single person, object, or event is called an **element** of the population.

The population of telephone book listees contains a **finite** number of elements; the population resulting from tossing the coin contains an **infinite** number.

A population is either concrete or conceptual. For example, the population of telephone book listees is **concrete**—given sufficient time you could contact each person because the number of elements is finite and the population is well defined.

The population of heads and tails is **conceptual**—try as you may, you cannot record all the results of tossing a coin for eternity. This population exists as an idea rather than as a material object.

A population could consist of all the students in a university (people), their cars (objects), or their pep rallies (events).

> The number or label used to represent an element of the population is called an **observation** or **datum.**

It is a measurable characteristic of the elements. The observation for students in a university might be their GPAs, their cars' gas mileage, or the number attending pep rallies. If 362 students attended the second pep rally, the observation for this event is 362 students. The selection of an appropriate population for an experiment is determined by the nature of the research questions that a researcher wants to answer as well as by such practical matters as the availability of population elements.

> A **sample** is a proper subset of a population.

That is, a sample can contain a single element or all but one of the population elements. For practical reasons—such as limited resources and time or because the population is infinite in size—most research is carried out with samples rather than with populations. It is assumed that the study of a sample will reveal something about the population. This leap of faith often appears to be justified, as when a laboratory technician analyzes a sample of a patient's blood or when an automobile manufacturer crash-tests a sample of bumpers. Occasionally, however, samples lead us astray. Later you'll see how and why.

Descriptive and Inferential Statistics

It is useful to divide statistical techniques into two broad categories: descriptive and inferential.

> **Descriptive statistics** are tools for depicting or summarizing data so that they can be more readily comprehended.

When we say that a player's lifetime batting average is .420 or when we determine that 51% of voters favor a presidential candidate, we are using descriptive statistics. A computer printout listing the Scholastic Aptitude Test (SAT) scores of all college students in California would boggle our minds; however, a statement that their mean SAT score is 1094 would not. Large masses of data are difficult to comprehend. Descriptive statistics reduce data to some form, usually a number, that one can easily comprehend. I discuss a variety of descriptive statistics in the first half of this book.

It is usually impossible for researchers to observe all the elements in a population. Instead they observe a sample of elements and generalize from the sample to all the elements—a process called **induction** in which the researcher reasons from the particular facts or cases to draw general conclusions.

Researchers are aided in this process by **inferential statistics,** which are tools for inferring the properties of one or more populations by inspecting samples drawn from the populations.

Inferential statistics were developed to improve decision making in cases where successive observations exhibit some degree of variation although they are obtained under conditions that appear to be identical. The variation may be due to (1) the inherent variability in the phenomenon being observed or differences among participants, (2) errors of measurement, (3) undetected changes in conditions, or (4) a combination of these factors. In the behavioral sciences, health sciences, and education, differences in the past experiences and heredities of participants are the major stumbling blocks to inferring the properties of populations from observing samples.

Inferential statistics are useful for answering questions such as the following. A medical researcher wants to know whether a new drug will arrest the development of cancer in humans. It is impossible to administer the drug to the population of all cancer patients, but it is possible to administer the drug to a sample. The medical researcher would probably attempt to control attitudinal and other extraneous factors by administering an inert druglike substance, a *placebo*, to half the sample and the new drug to the other half. Consider two possible outcomes of the experiment. In one outcome, the remission of cancer occurs in 100% of the sample receiving the new drug and in only 8% of those receiving the placebo. The difference, 100% versus 8%, between the drug and placebo samples is dramatic. The medical researcher would probably conclude without the benefit of inferential statistics that if the drug had been administered to the population of all cancer victims, the remission rate would have been much higher than if the population had received the placebo. Consider now a different outcome. What if the remission rate were only 12% for the new drug and 8% for the placebo? Is the drug really more effective than the placebo? I know from years of conducting experiments that chance factors can produce a difference between two samples even though the samples are taken from the same population and receive identical treatments. Is the difference, 12% versus 8%, greater than would be expected by chance? Stated another way, if the experiment were repeated many, many times, could the medical researcher predict with confidence that over the long run the difference would favor the sample receiving the drug? This is the kind of question that can be answered using inferential statistics. I describe procedures for answering such questions in the second half of the book.

Random Sampling

Some samples provide a sound basis for drawing conclusions about populations; others do not. The difference lies in the method by which the samples are selected.

The method of drawing samples from a population such that every possible sample of a particular size has an equal chance of being selected is called **random sampling,** and the resulting samples are **random samples.**

People, when left to their own devices, find it virtually impossible to produce random samples. Consider the following experiment. One hundred people are asked to write down a random sample of four numbers from the first 20 positive integers. According to our definition of random sampling, samples containing the elements 1, 2, 3, 4 or 14, 16, 18, or 20, for example, should occur as frequently as any other sample of size four. It turns out that such samples are rarely produced. People avoid writing down samples with consecutive or equally spaced integers and attempt to produce samples that span the range from 1 to 20.

Sampling methods based on haphazard or purposeless choices, such as soliciting volunteers, using students enrolled in introductory psychology, or selecting every 10th person in an alphabetical listing of names, produce **nonrandom samples.** Such samples, unlike random samples, do not provide a sound basis for deducing the properties of populations. Hence, in this book, sampling refers to random sampling. A detailed discussion of random sampling in Chapter 8 must await the development of other basic concepts. At this point, I will simply illustrate several characteristics of random samples.

Consider a box containing 300 balls, each identified by a number stamped on its surface. Of the balls, 200 are red (R) and 100 are black (B). If you did not know the ratio of red to black balls, which is two to one (denoted by 2:1), you could estimate the ratio by drawing a random sample of balls from the box. You close your eyes, shake the box vigorously, reach in, withdraw a ball, note its color and number, and replace it. You do this six times and obtain the following sample: R_{102}, R_{75}, B_{39}, R_{62} B_{37}, R_{50}. The subscripts, 102, 75, and so on denote the numbers stamped on the balls. From this sample you would infer that the box contains more red than black balls—in fact, twice as many red balls. Suppose you drew four more samples, each time replacing the balls drawn, and obtained the following:

Sample 2	$R_{154}, B_{62}, R_{35}, R_{143}, R_4, R_{29}$
Sample 3	$R_{104}, B_{41}, B_{21}, R_{50}, R_{192}, R_{67}$
Sample 4	$B_{28}, B_{41}, R_{150}, B_{61}, R_{88}, R_{148}$
Sample 5	$R_{152}, R_{120}, B_{88}, R_{33}, R_{36}, B_5$

The results of the five random samples are summarized in Table 1.3-1.

This simple experiment illustrates several points about random samples. First, the elements obtained (and the ratio of red to black balls) differ from sample to sample. This is referred to as **sampling fluctuation** or **chance variability.** Second, the

TABLE 1.3-1 Outcomes of Drawing Five Random Samples

	Sample				
Color of Balls	*1*	*2*	*3*	*4*	*5*
Number of red balls	4	5	4	3	4
Number of black balls	2	1	2	3	2
Ratio of red to black	2:1	5:1	2:1	1:1	2:1

characteristics of a sample do not necessarily correspond to those in the population. It turns out, however, that the larger a random sample, the more likely it is to resemble closely the population. Hence, researchers prefer to work with large samples if it is economically feasible. Although there is no guarantee that large random samples will resemble the population, in the long run they are more likely to do so than small ones.

CHECK YOUR UNDERSTANDING OF SECTIONS 1.1 TO 1.3[3]

1. Users of statistics fall into four categories.
   a. List the categories.
   b. Considering your vocational goals, into which category do you fall? Why?
2. For each of the following statements, indicate (a) the population, (b) the element, and (c) the observation to be recorded.
   a. At least 50% of white women students in this university are ambivalent about having a career.
   b. Tequila Tech students are involved in more automobile accidents than other drivers in their age group.
   c. At least 23% of the homes in Chickasha, Oklahoma, have high-definition televisions.
   d. Students at Ginebra University who hold outside jobs have higher grade point averages than those who do not hold outside jobs.
   e. According to a recent Centers for Disease Control report, 1 of every 92 American men between the ages of 27 and 39 has the AIDS virus.
   f. According to the U.S. Department of Education, 49.5% of female high school students have performed a community service during the past two years.
3. What are the lower and upper limits on the size of a sample?
4. Indicate whether each of the following procedures would produce a random sample (R) or a nonrandom sample (NR) of students in an introductory psychology class.
   a. Write each student's name on a slip of paper, place the slips in a hat, shake the hat thoroughly, and draw out 10 names.
   b. Place the blindfolded instructor in the middle of a circle made up of all the class members. Have the instructor point to 10 people around the circle. The student nearest to where the instructor points becomes an element of the sample.
   c. For each student, flip a fair coin. If the coin lands heads, the student is in the sample.
   d. Line up the students from the tallest to the shortest. The 3rd, 5th, 7th, . . . , 21st students become members of the sample.

[3] Answers to the *Check Your Understanding* exercises are given in Appendix C. These exercises often contain multiple questions about a particular concept. If you have a good grasp of the concept, answering three or four questions about it may not be an efficient use of your time. If, however, your answer to a question is incorrect, reviewing the concept in the text and then answering several more questions dealing with the concept is advisable. One of the purposes of these exercises is to provide feedback about what you know and what you don't know.

5. Terms to remember:
   a. Statistics b. Population
   c. Element d. Observation (datum)
   e. Sample f. Descriptive statistics
   g. Induction h. Inferential statistics
   i. Random sample j. Nonrandom sample
   k. Sampling fluctuation (chance variability)

1.4 DESCRIBING CHARACTERISTICS BY NUMBERS

People, objects, and events have many distinguishable characteristics. Early in the design of an experiment, the researcher must make two key decisions: What characteristics should I measure? And how should I measure them? The answer to the first question is determined by the researcher's interests. Suppose a researcher is interested in comparing the SAT scores of men and women college students. College students differ in many ways: gender, age, SAT scores, major, hair color, family income, and so forth, but only two characteristics are of interest in this example: gender and SAT score. The researcher will measure these characteristics and ignore the others. The second question, concerning how the characteristics should be measured, is less straightforward. The issue here is how to assign numbers to people, objects, or events so that the numbers accurately reflect the characteristic you want to measure. In the process of examining this issue, I will discuss variables and constants and see how mathematicians classify variables.

Variables and Constants

A **variable** is a characteristic that can take on different values. A variable also is a symbol, often a letter toward the end of the alphabet, such as X or Y, that is used to stand for an unspecified element of a set.

The set of elements for which the variable stands is called the **range** of the variable, and each element of the range is called a **value.** When I assign to a variable one of the elements in its range, I say that the variable "takes" this value. For example, the variable of gender might take the value "women."

A **constant** is a characteristic that does not vary. A constant also is a symbol, often a letter toward the beginning of the alphabet, such as a, b, or c, whose range consists of a single element.

The ratio of the circumference of a circle to its diameter, denoted by π, is a constant because its range consists of the single value 3.1415926536 . . . .

Perspectives on Numbers

I noted that the selection of the characteristics to be measured is relatively straightforward and is determined by the researcher's interests. The second key decision—deciding how the characteristics should be measured or classified—is not as simple. For example, you could measure or classify the scholastic aptitude of seniors at Linden McKinley High by (1) assigning each student a label such as average, high average, or superior, based on his or her SAT score; (2) ranking or ordering students' SAT scores from highest to lowest and assigning each student the number of his or her rank; or (3) assigning each student her or his actual SAT score. Depending on the measuring scheme adopted, Jonathan Whiz would be designated, respectively, superior, 3, or 1480. The variable of political preference can be classified by assigning a unique symbol such as D or 1 to Democrats, I or 2 to independents, and R or 3 to Republicans.

The assignment of numbers or labels to characteristics of people, objects, or events and the accuracy of the representation are central concerns of researchers. This is not true for mathematicians. Mathematicians often manipulate symbols that are totally devoid of empirical meaning. They are interested in the formal properties of the systems they create; applications in the real world are often left to other specialists. Mathematicians and mathematical statisticians have laid the foundation for a vast collection of statistical tools. The researcher who uses these tools must decide whether a particular tool is appropriate for his or her research application and whether the numbers assigned to variables accurately represent the characteristics of interest. This division of interest between the developers and the users of statistics has led to two ways of thinking about numbers.

Classification of Variables in Mathematics

Mathematicians classify variables as qualitative or quantitative.

A **qualitative variable** is a symbol whose range consists of attributes or nonquantitative characteristics of people, objects, or events. For example, the letter X could represent gender (men, women), Y could represent race (Caucasian, African American, Asian, other), and Z could represent the grade in a course (A, B, C, D, or F).

The categories of a qualitative variable are (1) mutually exclusive (nonoverlapping), which implies that an element cannot be in more than one category, and (2) exhaustive, which implies that an element must be in one of the categories. The categories may or may not suggest an order or rank. For example, grades in a course—A, B, C, D, or F—clearly order academic achievement from highest to lowest, but no order is suggested by the categories for gender, race, religious preference, or blood type. Course grade is an example of an **ordered qualitative variable.** Gender, race, religious preference, and blood type are examples of **unordered qualitative variables.**

A **quantitative variable** is a symbol whose range consists of a count or a numerical measurement of a characteristic.

Quantitative variables can be discrete or continuous. A variable is **discrete** if its range can assume only a finite number of values or an infinite number of values that is countable. That is, the infinite number of values can be placed in a one-to-one correspondence with the counting or natural numbers. Family size is an example of a variable with a finite range. It can assume values 1, 2, 3, 4, and so on, but not 200, 8000, or any noninteger value such as 0.5 and 4.3. The rational numbers—numbers that can be expressed as the ratio of two integers, for example, 2/2, −2/3, or 7/4—illustrate countably infinite numbers. There is no largest number and no smallest number, and between, say, 1 and 2, an infinite number of rationals can be inserted, for example, 3/2, 4/3, 5/4. . . . Other examples of discrete quantitative variables are the number of parking tickets received, the number of trials required to learn a list of nonsense syllables, and one's score on a standardized achievement test. In each of these examples, the value assigned to the variable is obtained by counting, and the counting units—family members, parking tickets, learning trials, or achievement test items—are equivalent in arriving at the total count.

By contrast, a variable is **continuous** if its range is uncountably infinite. Such a range can be likened to points on a line that have no interruptions or intervening spaces between them. Examples of continuous variables are temperature in Bangor, Maine, during January, length of fish caught off the Florida Keys, and speed of cars on the New Jersey Turnpike. Although a variable is continuous, our measurement of it is by necessity discrete because of limitations in the measuring instrument. For example, the thermometer is usually calibrated in 1° steps, the ruler in 1/16 inch, and the speedometer in 1 mile per hour. Consequently, our measurement of continuous variables is always approximate. Discrete variables, on the other hand, can be measured exactly. A husband and wife with two children are a family of exactly four, but a temperature of 80°F can be any temperature between 79.5° and 80.5°F.

The classification scheme for variables is summarized in Table 1.4-1. It is useful to mathematicians and statisticians because the nature of the variable determines which mathematical tools can be used to solve problems and do derivations and proofs. Hence, the classification scheme is a convenience; it was not devised to mirror characteristics in the real world. When you use statistical methods to answer real-world questions, you must remember that the methods were developed to analyze numbers as

TABLE 1.4-1 Mathematicians' Classification of Variables

Type of Variable	*Characteristics*
Qualitative variable	Range consists of nonoverlapping and exhaustive categories that represent attributes or nonquantitative characteristics.
Unordered	Categories do not suggest an order or rank.
Ordered	Categories suggest an order or rank.
Quantitative variable	Range consists of a count or a numerical measurement of a characteristic.
Discrete	Range consists of only a finite number of values or an infinite number of values that is countable.
Continuous	Range consists of an uncountably infinite number of values.

numbers. If the numbers analyzed bear no relation to the characteristics in which you are interested, the statistical methods will yield answers that are meaningless.

Measuring Operations in the Behavioral Sciences, Health Sciences, and Education

Numbers are used for a variety of purposes, three of which are of particular interest to behavioral scientists, health scientists, and educators: (1) to serve as labels, (2) to indicate rank in a series, and (3) to represent quantity. For example, a football player is identified by the number 10 on his uniform, a team is ranked number two in the UPI poll, and the winning touchdown play covered 20 yards. Without thinking, you treat these numbers differently. It doesn't take a football fan to know that player 30 is not three times player 10 and that the number two team is not necessarily twice as good as the number four team, but a 20-yard touchdown play did indeed move the ball twice as far down the field as a 10-yard play. You intuitively treat the numbers differently because they involve different levels of measurement.

Measurement is the process of assigning numbers or labels to characteristics of people, objects, or events according to a set of rules.

You will see that the rules used to assign the numbers or labels determine the level of measurement. S. S. Stevens (1946), a behavioral scientist, identified four levels of measurement: nominal, ordinal, interval, and ratio.

Nominal Measurement

Nominal measurement is the simplest of the four levels. It consists of assigning elements to mutually exclusive and exhaustive *equivalence classes* so that those in the same class are considered to be equivalent to one another, whereas those in different classes are not equivalent. The classes are then denoted by a set of distinct labels. The set of labels constitutes a **nominal scale.**

The assignment of men to one equivalence class called "men" and women to the other called "women" is nominal measurement. The set of labels, "men" and "women," constitutes a nominal scale. Numbers can be used instead of words to identify the two classes, for example, 1 for women and 2 for men. Numbers used in this way are simply alternative labels for the equivalence classes. You could just as well have assigned the numbers 9 and 6, respectively, to women and men. The substitution of the number 9 for 1 and the number 6 for 2 is an example of a **one-to-one transformation.**[4] The numbers 9 and 6 are as useful for distinguishing between the equivalence classes as any other one-to-one transformation. The numbers in a

[4] A one-to-one transformation associates with each element in one set one and only one element in a second set and vice versa. For example, if one set is men's names {Jim, Chuck, Keith} and the second set is numbers {5, 12, 3}, each name can be paired with one and only one number. A one-to-one transformation could result in substituting 12 for Jim, 3 for Chuck, and 5 for Keith.

nominal scale could be added, subtracted, averaged, and so on, but the resulting numbers would tell us nothing about the equivalence classes represented by the numbers. For example, $1 + 2 = 3$ and $9 + 6 = 15$, but neither 3 nor 15 corresponds to any characteristic of men or women. This follows because we did not utilize the properties of size and order of numbers when we assigned them to the classes. The only property of numbers that we utilized is that 1 is distinct (different) from 2, 3. . . .Thus, the labels assigned to equivalence classes in nominal measurement have the property only of *distinctness*.

There are many examples of nominal scales in psychology and education, for example, Eysenck's four personality types (stable-extrovert, stable-introvert, unstable-extrovert, unstable-introvert), the primary taste qualities (sweet, sour, salty, bitter), and categories of psychoses (organic, functional). There is a correspondence between a nominal scale and the range of one of the mathematician's types of variables. The nominal scale corresponds to the range of an unordered qualitative variable.

Ordinal Measurement

Ordinal measurement consists of assigning elements to mutually exclusive and exhaustive equivalence classes that are ranked or ordered with respect to one another. The classes are then denoted by numbers or other ordered symbols, such as letters of the alphabet, that reflect the rank of the classes. The labels assigned to equivalence classes in ordinal measurement have the properties of *distinctness* and *order*. The set of labels constitutes an **ordinal scale.**

The labels used in ordinal measurement contain more information than those in nominal scales: both distinctness and order.

The ranking of political candidates with respect to voter appeal is an example of ordinal measurement. If candidate Jane is judged to have the greatest appeal, followed by Keith, Lewis, and then Marvin, I could assign Jane the number 1; Keith, 2; Lewis, 3; and Marvin, 4. I have no reason to believe that Keith, ranked second, is half as appealing to voters as Jane, or that the difference in appeal between Jane and Keith, 1 versus 2, is the same as the difference between Keith and Lewis, 2 versus 3. The numbers indicate rank order but not magnitude or difference in magnitude between classes. The numbers assigned to the equivalence classes can be subjected to any strictly increasing monotonic transformation. A **strictly increasing monotonic transformation** permits one to replace the original set of numbers with new numbers as long as the new numbers have the same order as the original numbers. For example, the set of ordered numbers 2, 16, 39, 40 would serve just as well as 1, 2, 3, 4 to rank the four candidates, because only the order and not the distance between any two numbers is important. Alternatively, I could assign the ordered letters of the alphabet to the candidates: *A* to Jane, *B* to Keith, *C* to Lewis, and *D* to Marvin. The transformations that can be applied to ordinal scales are more restrictive than those that can be applied to nominal scales. This follows because the labels in ordinal scales contain more information that needs to be preserved—both distinctness and order—than do the labels in nominal scales.

Some characteristics, such as people's heights, can be measured in several ways, for example, ranking from tallest to shortest or recording actual feet and inches. The latter procedure assigns numbers that represent the magnitudes of the equivalence classes and therefore has several advantages over ordinal measurement, as you shall see later. For the moment, simply note that ordinal measurement is most often used when it is difficult or impossible to apply more refined measuring procedures. For example, it is difficult to precisely measure the tastiness of three pizzas or the leadership qualities of four political candidates. However, it is not too difficult to rank-order pizzas with respect to tastiness or candidates with respect to leadership qualities.

Numerous examples of ordinal scales can be found in the behavioral sciences, health sciences, and education, for example, classification of mentally subnormal children (borderline, educable, trainable, profoundly retarded) and professorial rank (instructor, assistant professor, associate professor, professor). Such ordinal scales correspond, in the language of the mathematician, to the range of an ordered qualitative variable.

Interval Measurement

The numbers assigned in interval measurement contain much more information than the labels used in nominal and ordinal measurement.

In **interval measurement,** the numbers assigned to equivalence classes have the properties of distinctness and order; in addition, equal differences between numbers reflect equal magnitude differences between the corresponding classes. The measurement procedure consists of defining a unit of measurement, such as a calendar year or 1°F, and determining the number of units required to represent the difference between equivalence classes. The set of numbers assigned to the equivalence classes constitutes an **interval scale.**

In our measurement of calendar time, the same amount of time elapsed between 1970 and 1971 as between 1971 and 1972, and, similarly, the temperature difference between 70° and 75°F is the same as that between 80° and 85°F. A given numerical interval, say 1 year or 5°F, represents the same difference in the characteristic measured, irrespective of the location of that interval along the measurement scale. In other words, numerically equal distances along the measurement continuum represent empirically equal differences among the corresponding equivalence classes—that is, the measured characteristic.

Because the units of measurement along interval scales are empirically equal, it is meaningful to perform most arithmetic operations on the numbers. For example, I can say that the difference between 80° and 60°F is twice as great as that between 60° and 50°F. That is, the ratio of intervals $(80° - 60°F)/(60° - 50°F) = 2$ has meaning with respect to temperature. However, not all arithmetic operations are permissible because the starting point or origin of an interval scale is always arbitrarily defined and does not correspond to an absence of the measured characteristic. In the case of the Fahrenheit scale, 0°F corresponds to the temperature produced by mixing equal quantities by weight of snow and salt. This 0 does not indicate an absence of molecular action and hence an absence of heat. Therefore, although 80°F/40°F = 2, I cannot say that 80°F is twice as hot as 40°F. The ratio 80°F/40°F = 2 is uninterpretable because the zero point on the scale, 0°F, does not correspond to the absence of temperature. The same interpretation problem occurs for calendar time, which is measured from the birth of Christ, and altitude, which is measured from sea level.

The numbers in an interval scale can be subjected to any positive linear transformation. A **positive linear transformation** of a variable, say X, consists of multiplying X by a positive constant b and adding a constant a to the product. That is, a transformed value, X', is given by $X' = a + bX$. For example, degrees Fahrenheit, F, can be transformed into degrees Celsius, C, by means of the positive linear transformation

$$X' = a + bX$$

$$C = \frac{5}{9}(-32) + \frac{5}{9}\text{F},$$

where $X' = \text{C}$, $a = 5/9\,(-32)$, $b = 5/9$, and $X = \text{F}$. Although the variable represented by an interval scale may be continuous, our measurement of it is always discrete because measuring instruments are calibrated in discrete steps. Thus, in practice an interval scale corresponds to the range of a discrete quantitative variable.

Ratio Measurement

The numbers assigned in ratio measurement contain the most information.

In **ratio measurement,** the numbers assigned to equivalence classes have the properties of distinctness, order, and equivalence of intervals; in addition, the origin of the scale represents the absence of the measured characteristic. The set of numbers assigned to the equivalence classes constitutes a **ratio scale.**

Ratio scales have all the properties of interval scales plus an absolute zero. Most scales in the physical sciences are ratio scales—height in inches, weight in pounds, temperature on the Kelvin scale, and elapsed time such as the age of an object.

Not only is the difference between 5 and 6 inches the same distance as that between 10 and 11 inches, but also an object that is 10 inches long is twice as long as an object that is 5 inches long. Ratio scales permit you to make meaningful statements about the ratio of the numbers assigned to the two objects, for example, 10 inches/5 inches = 2; hence 10 inches is twice as long as 5 inches. The properties of a ratio scale mentioned in the previous paragraph permit you to perform all arithmetic operations on the numbers. However, the only transformation of a ratio scale that preserves these properties is **multiplication by a positive constant:** $bX = X'$, where b is a positive number, X is the original value, and X' is the transformed value. For example, I can transform inches into centimeters by multiplying inches by the constant $b = 2.54$: 10 inches is equal to

$$(2.54)(10 \text{ in.}) = 25.4 \text{ cm}$$

and 5 inches is equal to

$$(2.54)(5 \text{ in.}) = 12.7 \text{ cm}$$

Ten inches is twice as long as 5 inches and, similarly, 25.4 centimeters is twice as long as 12.7 centimeters. As I move from measurement in which the labels contain the least information (nominal scales) to those containing more information (ordinal, interval, and ratio scales), more and more constraints are placed on the transformations that can be meaningfully applied. This occurs because the numbers in ordinal, interval, and ratio scales contain more information that can be altered or destroyed by a transformation. In practice, a ratio scale, like the interval scale, corresponds to the range of a discrete quantitative variable. The major characteristics of the four scales are summarized in Table 1.4-2.

TABLE 1.4-2 Overview of Levels of Measurement

Level of Measurement	*Characteristics*
Nominal	Symbols serve as labels for mutually exclusive and exhaustive equivalence classes. The symbols have the property of distinctness. *Appropriate transformation*: any one-to-one substitution. *Corresponds to*: range of an unordered qualitative variable. *Examples*: gender, eye color, racial origin, personality types, and primary taste qualities.
Ordinal	Ordered symbols, usually numbers, indicate rank order of equivalence classes. The symbols have the properties of distinctness and order. The size of differences between ordered symbols provides no information about differences between equivalence classes. *Appropriate transformation*: monotonic. *Corresponds to*: range of an ordered qualitative variable. *Examples*: military rank, classification of mentally retarded children, rank in high school, and a supervisor's ranking of employees.
Interval[a]	Equal differences among numbers reflect equal magnitude differences among equivalence classes, but the origin or starting point of the scale is arbitrarily determined. Numbers have the properties of distinctness, order, and equivalence of intervals. *Appropriate transformation*: positive linear. *Corresponds to*: range of a discrete quantitative variable. *Examples*: Fahrenheit and Celsius temperature scales, calendar time, and altitude.
Ratio[a]	All the properties of interval scales apply, and, the origin of the scale reflects the absence of the measured characteristic. *Appropriate transformation*: multiplication by a positive constant. *Corresponds to*: range of a discrete quantitative variable. *Examples*: height, weight, Kelvin temperature scale, and measures of elapsed time.

[a] These two levels are sometimes referred to collectively as **metric measurement** or **numerical measurement.**

Implications of the Two Ways of Thinking about Numbers

This chapter has described two ways of thinking about numbers: one reflects the concerns of mathematicians, and the other reflects the concerns of behavioral scientists, health scientists, and educators. We have developed statistical methods for analyzing numbers as numbers, whether or not the numbers are true measures of some characteristic. If the assumptions associated with the statistical methods are fulfilled, they will produce answers that are formally correct as numbers. This is true regardless of the degree of correspondence between the numbers and the characteristic they represent. The problem comes in translating statistical results into statements about the real world. If numbers representing a nominal scale are manipulated arithmetically, the result will be numbers that are numerically correct but uninterpretable. If nonsense is put into the equation, nonsense indeed will come out.

Most researchers are very sensitive to the potential pitfalls associated with interpreting numbers produced by statistical procedures—and rightfully so. Some authors have even gone so far as to prescribe the statistical procedures that can be used with each measurement scale.[5] Except in the physical sciences, few scales have equal intervals, so the number of statistical techniques on the approved list is relatively small. However, this position fails to recognize that the measurement of many variables in the behavioral sciences and education lies somewhere between the ordinal and interval levels. The IQ scale is a good example. Most psychologists and educators agree that the 10-point difference between IQs of 100 and 110 represents a slightly smaller intellectual difference than the 10-point difference between IQs of 130 and 140. Although the 10-point differences do not represent identical intellectual differences, the intellectual differences are believed to be similar. Hence, IQ scores contain more information than ordinal scales but less than interval scales.

Another example of a measurement scale that is between the ordinal and interval levels is the attitude rating scale: strongly disagree $= -2$, disagree $= -1$, neutral $= 0$, agree $= 1$, strongly agree $= 2$. The numbers $-2, -1, 0, 1, 2$ contain ordinal information. However, it is unlikely that the actual difference in attitudes between 0 and 1, for example, is identical to the difference between 1 and 2. But the difference in attitudes between 0 and 1 is probably similar to the difference between 1 and 2. Thus, the five numbers along the attitude scale do contain some information about the magnitude differences in attitudes.

Should we avoid performing arithmetic operations on scores when the measurement is between the ordinal and interval levels? Researchers have heatedly debated this question. We cannot look to mathematicians and statisticians for answers because the question is outside their province. The answer must come from users of statistics who are acquainted with the problems of translating numerical answers into statements about the real world. An examination of the professional literature reveals that most experts in the behavioral sciences, health sciences, and education do apply arithmetic operations to numbers even though the measurement is somewhere between the ordinal and interval levels. Further, they interpret

[5] Examples can be found in Senders (1958), Siegel (1956), and Stevens (1946, 1951).

the results as if the size of a difference between the numbers reflects something about the size of a difference in the measured characteristics. Apparently, experts prefer to utilize whatever magnitude information the numbers contain, even though differences among the numbers only approximate the true magnitude differences.

If a researcher believes that any transformation of a set of numbers that preserves the order of the original numbers adequately represents the equivalence classes, the numbers contain no magnitude information, and they should not be treated as though they do. In the final analysis, it is the researcher, the person most familiar with the data, who must decide how much information the numbers contain.

Some Subtle Problems in Interpreting Numbers

The preceding discussion has emphasized the importance of avoiding interpretation errors by being sensitive to the degree of correspondence between a set of numbers and the characteristic they represent. Consider now some not-so-obvious interpretation problems that occur when a test has an arbitrary zero point. Suppose that on a standardized arithmetic-achievement test, Mortimer received a score of 0; Dude, a score of 30; and Reginald, a score of 60. Can you conclude that Mortimer knows nothing about arithmetic? Obviously not; a score of 0 means that he couldn't answer any questions on the test, but easier questions may exist that he could answer. Achievement tests, as well as many other tests, have arbitrary rather than absolute zero points and therefore fall short of ratio measurement. It follows that although Reginald's score of 60 is twice as high as Dude's 30, Reginald's arithmetic achievement isn't necessarily twice Dude's.

The interpretation problem that results from a lack of equal intervals is subtler. Suppose I compare the effectiveness of two methods of teaching arithmetic. Students in a class using method A gained an average of 10 points; those in a class using method B gained an average of 7 points. The results seem straightforward—on the average, students using method A gained more points than those using method B. But suppose that at the beginning of the experiment the two classes were not equal in arithmetic achievement. Let the average score for class A be 50 and the average score for class B be 80. Is it possible that a 7-point change from 80 to 87 represents more improvement in arithmetic achievement than a 10-point change from 50 to 60? Unless I know that, say, a 10-point change anywhere on the measurement scale represents the same empirical change, the interpretation of the experiment is equivocal. The greater the difference between the classes' initial average achievement scores, the greater the interpretation problem.

Consider finally the interpretation problem that occurs when a test does not have enough difficult items to adequately differentiate among high-scoring participants. Suppose that two individuals make the top score of 60. For one participant, this may represent maximum capability, but the other person may be capable of a much higher performance. The measuring instrument is simply incapable of showing it. Because of the limitations of the measuring instrument, it would be incorrect to conclude that the two individuals are equal in the characteristic measured.

Because numbers do not always mean what they appear to mean, they must be carefully scrutinized. The key principle that runs throughout this section is that a researcher must be guided by two sets of rules. When the tools of statistics are used, the mathematician's and statistician's rules must be followed. When the numbers are interpreted as statements about the real world, the behavioral scientist's measurement rules must be followed.

CHECK YOUR UNDERSTANDING OF SECTION 1.4

6. Ignoring for the moment the limitations of measuring instruments, classify measures of the following according to the mathematician's scheme (unordered qualitative, U; ordered qualitative, O; discrete quantitative, D; continuous quantitative, C).
   a. Size of family
   b. Race
   c. Paper and pencil test of marital compatibility
   d. Seeding of tennis players in a tournament
7. Because of the limitations of measuring instruments, measurement of some variables is of necessity approximate. Classify the variables in Exercise 6 according to whether our measurement is exact (E) or approximate (A).
8. Reclassify the variables in Exercise 6 according to the mathematician's scheme, taking into account limitations in our ability to measure some of the variables.
9. Classify the variables in Exercise 6 with respect to the level of measurement, taking into account limitations in our ability to measure some of the variables.
10. For each level of measurement, indicate the appropriate transformation that can be performed on the numbers.
11. Four kinds of transformations are described in this section. For each level of measurement, list all of the kinds of transformations that can be performed without altering the information contained in the original measurements.
12. What level of measurement is most often achieved (a) in the physical sciences and (b) in the behavioral sciences and education?
13. A score of 0 on an achievement test does not necessarily mean that the individual knows nothing about the subject. Explain.
14. Suppose that achievement test scores for a control group increased from 62 to 65, and those for the experimental group increased from 68 to 74. What must be true to conclude unequivocally that the experimental group improved twice as much as the control group?
15. Terms to remember:
    a. Variable
    b. Range of variable
    c. Value of variable
    d. Constant
    e. Qualitative variable
    f. Quantitative variable
    g. Discrete variable
    h. Continuous variable
    i. Measurement
    j. Nominal scale
    k. One-to-one transformation
    l. Ordinal scale
    m. Monotonic transformation
    n. Interval scale
    o. Positive linear transformation
    p. Ratio scale