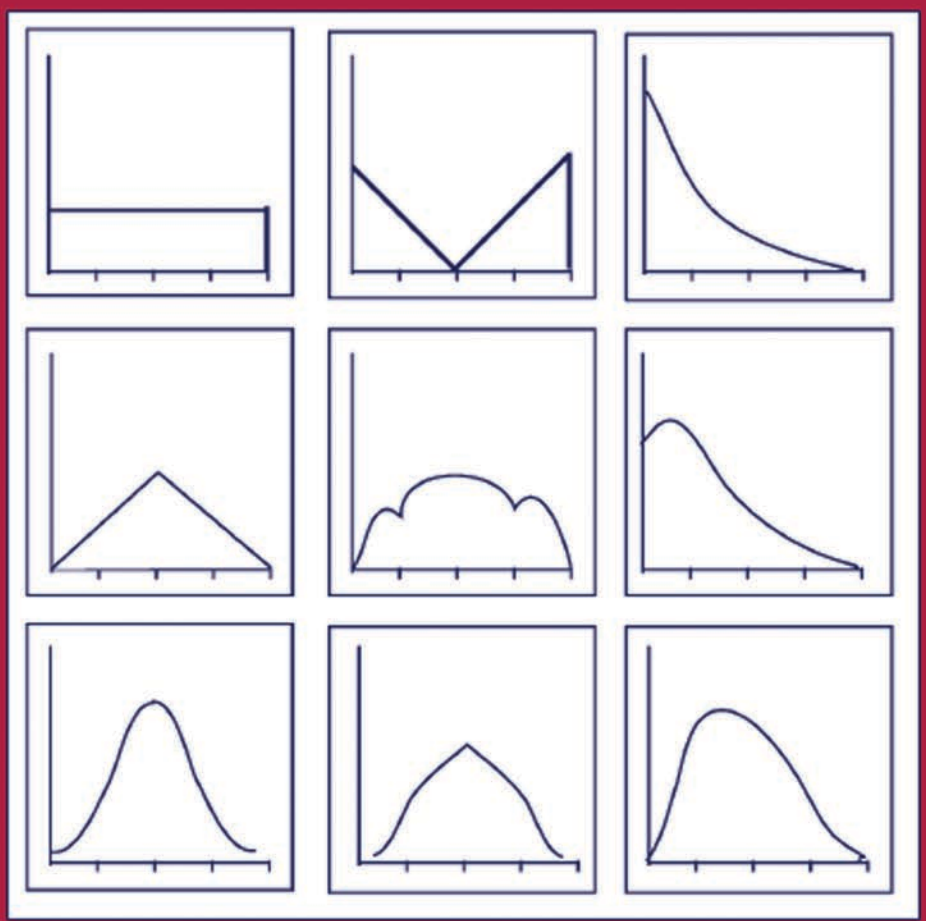
LEARNING FROM

DATA

AN INTRODUCTION TO STATISTICAL REASONING

THIRD EDITION



ARTHUR M. GLENBERG
MATTHEW E. ANDRZEJEWSKI

LEARNING FROM DATA

AN INTRODUCTION TO STATISTICAL REASONING

THIRD EDITION

Supplementary Resources Disclaimer

Additional resources were previously made available for this title on CD. However, as CD has become a less accessible format, all resources have been moved to a more convenient online download option.

You can find these resources available here: www.routledge.com/9780805849219

Please note: Where this title mentions the associated disc, please use the downloadable resources instead.

LEARNING FROM DATA

AN INTRODUCTION TO STATISTICAL REASONING

THIRD EDITION

ARTHUR M. GLENBERG
MATTHEW E. ANDRZEJEWSKI

Psychology Press
Taylor & Francis Group
NEW YORK AND LONDON

Psychology Press
Taylor & Francis Group
711 Third Avenue
New York, NY 10017

Psychology Press
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

Psychology Press is an imprint of Taylor & Francis Group, an informa business

International Standard Book Number-13: 978-0-8058-4921-9 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Glenberg, Arthur M.
Learning from data : an introduction to statistical reasoning / Arthur M. Glenberg and Matthew E. Andrzejewski. -- 3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-8058-4921-9 (alk. paper)
1. Statistics. I. Andrzejewski, Matthew E. II. Title.

HA29.G57 2008
001.4'22--dc22 2007022035

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

Contents

Chapter **22**

Preface

Statistics is a difficult subject. There is a lot to learn, and much of it involves new thinking. As the title implies, *Learning From Data: An Introduction to Statistical Reasoning* teaches you a new way of thinking about and learning about the world. Our goal is to put readers in a good position to understand psychological data and their limitations. Another more important goal is to evaluate data that affect all aspects of life—psychological, social, educational, political, and economic—to better prepare readers to question and to challenge. Yet another goal is to help readers retain the material. Psychologists have developed (from data) techniques that facilitate learning and comprehension, and we have incorporated three of these techniques into the book.

First, we have devoted extra attention to explaining difficult-to-understand concepts in detail. For example, some textbooks attempt to combine important concepts such as sampling distributions, hypothesis testing, power, and parameter estimation in one chapter. In this book, each concept has its own chapter. Yes, this means more reading, but it also means greater understanding.

Second, the book uses repetition extensively to help students learn and retain concepts. There are multiple fully explained examples of each major procedure. Many concepts (for example, power, Type I errors) are repeated from chapter to chapter. The problem sets at the ends of most chapters require students to apply principles introduced in earlier chapters.

The third major learning aid is the use of a consistent schema (the six-step procedure) for describing all statistical tests from the simplest to the most complex. The schema provides a valuable heuristic for learning from data. Students learn (1) to consider the assumptions of a statistical test, (2) to generate null and alternative hypotheses, (3) to choose an appropriate sampling distribution, (4) to set a significance criterion and generate a decision rule, (5) to compute the statistic of interest, and (6) to draw conclusions. Learning the schema at an early stage (in Chapter 8) will ease the way through Chapters 11 through 22, in which the schema is applied to many different situations. This schema also provides a convenient summary for each hypothesis-testing procedure. A table with a summary schema is included in the last section of each chapter containing the hypothesis-testing procedure. Inside the front cover of the book is a "Statistical Selection Guide" to further assist students in determining which statistical test is most appropriate for the situation.

About the book

There are many aspects to *Learning From Data* that differentiate it from other statistics textbooks. In addition to the three teaching/learning methods mentioned earlier, the content and organization of the book may be quite different from what students are used to. First, nonparametric statistical tests are integrated into the chapters in which analogous parametric tests are described. With this organization, students can better appreciate the situations in which particular tests apply. In fact, throughout the book there is an emphasis on practicing how to choose the best statistical procedure. The choice of the procedure is discussed in examples, and students are required to make the correct choice as they solve the problems at the end of the chapter. The endpapers of the book provide guidelines for choosing procedures.

Second, the initial parts of the chapters on regression (Chapter 20) and correlation (Chapter 21) are self-contained sections that include discussions of regression and correlation as descriptive procedures. Instructors may present these topics along with other descriptive statistics or delay their introduction until later in the course.

Third, the book contains two independent treatments of power. The major treatment begins in Chapter 9 with graphical illustrations of how power changes under the influence of such factors as the significance level and sample size. The chapter also introduces formulas for computing power and estimating sample size needed to obtain a particular level of power. These formulas are repeated and generalized for many of the statistical procedures discussed in later chapters. Often, however, there may not be enough time for an extensive treatment of power. In that case, instructors can choose to treat power less extensively and omit Chapter 9 (and the relevant formulas in the other chapters). This less extensive treatment of power is part of each new inferential procedure. It consists of a nonmathematical discussion of how power can be enhanced for that particular procedure.

Fourth, factorial designs, interactions, and the ANOVA are explained in greater detail than in most introductory textbooks. Our goal is to give students enough information so that they will be able to understand the statistics used in many professional journal articles. Of course, it would be foolish for the authors of any introductory textbook to try to cover the statistical analyses of complex situations. Instead, Chapter 18 discusses how two-factor and three-factor factorial experiments are designed, and how to interpret main effects and two-factor and three-factor interactions. Chapter 19 presents a description of computational procedures for the relatively simple two-factor, independent sample ANOVA.

Last, but most important to us, is Chapter 14, "Random Sampling, Random Assignment, and Causality." A major reason for writing the first two editions of this book was to address the issues discussed in this chapter. All of us who teach statistics courses and conduct research have been struck by the incongruity between what we practice and what we preach. When we teach a statistics course, we emphasize random sampling from populations. But in most experiments we do no such thing. Instead, we use some form of random assignment to conditions. How can we perform statistical analyses of our experiments when we have ignored the most important assumption of the statistical tests? In Chapter 14, we develop a rationale for this behavior, but the rationale extracts severe payment by placing restrictions on the interpretation of the results when random assignment is used instead of random sampling.

New to the third edition

In addition to the features already described, there are a number of new features. First, the third edition of *Learning From Data* is designed to be used seamlessly with Excel™. Unlike other texts that concentrate on statistical software, we choose to focus on Excel, a spreadsheet program. Recent versions of statistical programs produce output that are far more complicated than needed for the undergraduate level. The output from Excel is straightforward; however, the statistical tools available are not complete. Thus, we have written an Add-in ("LFD3 Data Analysis Add-in") for Excel so all the analyses presented in the book can be conducted in Excel. Excel is widely available and can also be used as a database, data manager, and graphics program; experience with these functions may provide a valuable set of skills for undergraduates in a number of professions, including psychology. Thus, files containing all the data used in the book are provided on a companion CD in Excel format. However, because other programs are still widely used, text-based files are also available for use in other statistical programs, like SPSS™, SAS™, and Systat™.

Second, the book attempts to capture the student's interest by focusing on what can be learned from a statistical analysis, not just on how it is done. This is most apparent in the treatment of hypothesis testing. Using the six-step schema, the last step in hypothesis-testing is described as deciding whether to reject the null hypothesis *and then* concluding what that decision implies about the world and what the implications for future action might be. Another way that the book attempts to capture the student's interest is by continually referring back to two real data sets. These data sets are intrinsically interesting and save time because new experimental scenarios do not need to be continually introduced. The first data set on the effectiveness of Zyban® and nicotine-replacement gum on smoking comes from Dr. Timothy Baker. Data from 608 participants are included on the companion CD. The second data set on the effects of having a child on marriage comes from Dr. Janet Hyde and Dr. Marilyn Essex. The data from 244 families are also included on the companion CD. Data from these studies are used throughout the book in illustrating important concepts. The fact that these are real data sets strikes a chord with students that statistics plays an important role in *Learning From Data*.

Finally, we have provided instructors with substantial resources. To begin with, we have added approximately 20 new problems to the end-of-chapter exercises and provided many more on the companion CD. Included on the instructor CD are sample test questions, exercises, and sample data sets. We have also generated Powerpoint® lectures for each chapter for instructors to use or edit, as they choose. There are a number of very useful graphics and illustrations that mirror the ones in the book. There are also fun, interactive exercises/demonstrations and tools that we have found useful (for example, data generation algorithms, Gaussian random number generators, etc.). As additional items become available, our Web site (www.LFD3.net) will provide users of the textbook access to them.

MANy thANks

Many people have contributed to this book. We thank our students and colleagues at the University of Wisconsin–Madison and those instructors who used the first two editions and

provided valuable comments. We also thank Laura D. Goodwin (University of Colorado, Denver), Richard E. Zinbarg (Northwestern University), Daniel S. Levine (University of Texas, Arlington), and Randall De Pry (University of Colorado, Colorado Springs) for their valuable reviews of many of the chapters and of the proposal for a third edition of the book. AMG thanks his instructors at the University of Michigan and Miami University. MEA thanks his instructors at Temple University, especially Ralph Rosnow, Alan Sockloff, and Phil Bersh. Thanks are due to the editorial and production staffs at Lawrence Erlbaum Associates, who tolerated delay after delay. Finally, thanks to Mina and Anna for their love and support.

Arthur M. Glenberg

Matthew e . Andrzejewski

Chapter 1

Why Statistics?

There are many ways to learn about the world and the people who populate it. Learning can result from critical thinking, asking an authority, or even from a religious experience. However, collecting data (that is, measuring observations) is the surest way to learn about how the world really is.

Unfortunately, data in the behavioral sciences are messy. Initial examination of data reveals no clear facts about the world. Instead, the data appear to be nothing but an incoherent jumble of numbers. To learn about the world from data, you must first learn how to make sense out of data, and that is what this textbook will teach you. *Statistical procedures are tools for learning about the world by learning from data.*

To help you to understand the power and usefulness of statistical procedures, we will explore two real (and important!) data sets throughout the course of the book. One of the data sets is courtesy of Professor Timothy Baker at the University of Wisconsin Center for Tobacco Research and Intervention (which we will call the Smoking Study). The data were collected to investigate several questions about smoking, addiction, withdrawal, and how best to quit smoking. The data set consists of a sample of 608 people who wanted to quit smoking. These people were randomly assigned (see Chapter 14 for the benefits of random assignment) to three groups. The participants in one group were given the drug bupropion

SR (Zyban) along with nicotine replacement gum. In a second group, the participants were given the bupropion along with a placebo gum that did not contain any active ingredients. The final group received both a placebo drug and a placebo gum. The major question of interest is whether people are more successful in quitting smoking when the the active gum is added to the bupropion. These data are exciting for a couple of reasons. First, given the tremendous social cost of cigarette smoking, we as a society need to figure out how to help people overcome this addiction, and these data do just that. Second, the study included measurements of about 30 other variables to help answer ancillary questions. For example, there are data on how long people have smoked and how much they smoked; data on health factors and drug use; and demographic data such as gender, ethnicity, age, education, and height. These variables are described more fully within the Excel and SPSS data files on the CD that comes with this book and in Appendix A. The statistical tools you will learn about will give you the opportunity to explore these data to the fullest extent possible. You can ask important questions—some that may never have been asked before—such as whether drug use affects people's ability to quite smoking, and you can get the answers. In addition, these data will be used to illustrate various statistical procedures, and they will be used in the end-of-chapter exercises.

The second data set is courtesy of Professors Janet Hyde and Marilyn Essex of the University of Wisconsin–Madison. The data set is a subset of the data from the Wisconsin Maternity Leave and Health Project and the Wisconsin Study of Families and Work (we will refer to it as the Maternity Study). This project was designed to answer questions about how having a baby affects family dynamics such as marital satisfaction, and how various factors affect child development. The data set consists of measurements of 26 variables for 244 families. Some of these variables are demographic, such as age, education, and family income. Marital satisfaction was measured separately for mothers and fathers both before the child was born (during the 5th month of pregnancy) and at three times after the birth (1, 4, and 12 months postpartum). There are also data on how much the mother worked outside the house and how equally household tasks were divided among the mothers and fathers. Finally, there are eight measures of the quality of mother–child interactions at 12 months after birth, and three measures of child temperament (for example, hyperactivity) measured when the child was 4.5 years old. These variables are described more fully on the CD that comes with this book and in Appendix B. As with the smoking data, you are free to use these data to answer important questions, such as whether the amount of time that a mother works affects child development.

This chapter introduces a number of topics that are basic to statistical analyses. We begin with a discussion of variability, the cause of messy data, and move on to the distinctions between population and sample, descriptive and inferential statistics, and types of measurement found in the behavioral sciences.

VAr i AbiLit y

The first step in learning how to learn from data is to understand why data are messy. A concrete example is useful. Consider the CESD (Center for Epidemiologic Studies Depression) scores from the Smoking Study (see Appendix A). Each participant rated 20 questions

such as "I felt lonely" using a rating of 0 (rarely or none of the time during the past week) to 3 (most of the time during the past week). The score is the sum of the ratings for the participant. For the 601 participants for whom we have CESD scores, the scores range from 0 to 23. About a quarter of the scores are below 2, but another quarter are above 9. These data are messy in the sense that the scores are very different from one another.

> **Variability** is the statistical term for the degree to which scores (such as the depression scores) differ from one another.

Chapter 3 presents statistical procedures for precisely measuring the variability in a set of scores. For now, only an intuitive understanding of variability is needed. When the scores differ from one another by quite a lot (such as the depression scores), variability is high. When the scores have similar values, variability is low. When all the scores are the same, there is no variability.

s ources of Variability

It is easy enough to see that the CESD data are variable, but why are they variable? In general, variability arises from several sources. One source of variability is individual differences: Some smokers are more depressed than others; some have difficulty reading and understanding the items on the test; some smokers' answers on the inventory are more honest than the answers of other smokers. There are as many potential sources of variability due to individual differences as there are reasons for why one person differs from another in intelligence, personality, performance, and physical characteristics.

Another source of variability is the procedure used in collecting the data. Perhaps some of the smokers were more rushed than others; perhaps some were tested at the end of the day and were more tired than others. Any change in the procedures used for collecting the data can introduce variability. Finally, some variability may be due to conditions imposed on the participants, such as whether they are taking the placebo gum.

Variables and Constants

Variability does not occur only in textbook examples; it is characteristic of all data in the behavioral sciences. Whenever a behavioral scientist collects data, whether on the incidence of depression, the effectiveness of a psychotherapeutic technique, or the reaction time to respond to a stimulus, the data will be variable; that is, not all the scores collected will be the same. In fact, because data are variable, collecting data is sometimes referred to as measuring a variable (or a random variable).

> **A variable** is a measurement that changes from one observation to the next.

CESD is a variable because it changes from one smoker (observation) to the next. "Effectiveness of a psychotherapeutic technique" is another example of a variable, because a given technique will be more effective for some people than for others.

Variables should be distinguished from constants.

> **Constants** are measurements that stay the same from one observation to the next.

The boiling point of pure water at sea level is an example of a constant. It is always 100 degrees Centigrade. Whether you use a little water or a lot of water, whether the water is encouraged to boil faster or not, no matter who is making the observation (as long as the observer is careful!), the water always boils at the same temperature. Another constant is Newton's gravitational constant, the rate of acceleration of an object in a gravitational field (whether the object is large or small, solid or liquid, and so on).

Many of the observations made in the physical sciences are observations of constants. Because of this, it is easy for the beginning student in the physical sciences to learn from data. A single careful observation of a constant tells the whole story.

You may be surprised to learn that there is not one constant in all of the behavioral sciences. There is no such thing as *the* effectiveness of a psychotherapeutic technique, or *the* depression score, because measurements of these variables change from person to person. In fact, because what is known in the behavioral sciences is always based on measuring variables, even the beginning student must have some familiarity with statistical procedures to appreciate the body of knowledge that comprises the behavioral sciences and the limitations inherent in that body of knowledge. In case you were wondering, this is why you are taking an introductory statistics course, and your friends majoring in the physical sciences are not.

The concept of variability is absolutely basic to statistical reasoning, and it will motivate all discussions of learning from data. In fact, the remainder of this chapter introduces concepts that have been developed to help cope with variability.

Populations and Samples

The psychologists studying addiction *might* be interested in the CESD scores of the specific smokers from whom they collected data. However, it is likely that they are interested in more than just those individuals. For example, they may be interested in the incidence of depression among all smokers in Wisconsin, or all smokers in the United States, or even all smokers in the world. Because depression is a variable that changes from person to person, the specific observations cannot reveal everything the researchers might want to know about all of these depression scores.

Statistical Populations

> A statistical **population** is a collection or set of measurements of a variable that share some common characteristic.

One example of a population is the set of CESD scores of all smokers in Wisconsin. These scores are measurements of a variable (CESD), and they have the common characteristic of being from a particular group of people: smokers in Wisconsin. A different statistical

population consists of the CESD scores for smokers in the United States. And, a very different population consists of the marital satisfaction scores for new mothers who work full-time outside of the home. The point is that you should not think of statistical populations as groups of people, such as the people in the United States. There is only one population of people for the United States, but there are an infinite number of statistical populations depending on what variables are measured (for example, CESD or marital satisfaction), and how those scores might be grouped (for example, smokers or working mothers).

Thinking of statistical populations as sets of measurements may appear cold and unfeeling. Nonetheless, thinking this way has a tremendous advantage in that it facilitates the application of the same statistical procedure to a variety of populations. Instead of having to learn one technique for analyzing and learning from depression scores, and another technique for analyzing IQ scores, and yet another for analyzing errors rats make in learning mazes, many of the same procedures can be applied in all of these cases. In every case we are dealing (statistically) with the same stuff, a set of measurements.

Unfortunately, thinking of statistical populations as sets of numbers can cause some people to become bored and lose interest in the enterprise. The way to counter this boredom is to remember that the statistical procedures are operating on numbers that have meaning: The numbers are scores that represent something interesting about the world (for example, the incidence of depression in smokers). As you read through this book, think about applying your new knowledge to problems that are of interest to you, and not just as manipulation of numbers.

t he Problem of Large Populations

Some statistical populations consist of a manageable number of scores. Usually, however, statistical populations are very large. For example, there are potentially millions of CESD scores of smokers. When dealing with large populations, it is difficult and time consuming to actually collect all of the scores in the population. Sometimes, for ethical reasons, all the scores in the population cannot be obtained. For example, suppose that a medical researcher believes that she has discovered a drug that safely and effectively reduces high blood pressure. One way to determine the drug's effectiveness is to administer it to all people suffering from high blood pressure and then to measure their blood pressures. (The population of interest consists of the blood pressure scores of people suffering from high blood pressure who have taken the new drug.) Clearly, this would be time consuming and expensive. It would also be very unethical. After all, what if the medical researcher were wrong, and the drug did more harm than good? Also, even with a great national effort, not all the scores could be collected, because some of the people would die before they took the drug, others would have their blood pressures lowered by other drugs, and others would develop high blood pressure over the course of data collection.

We appear to have run across a problem. Usually, we are not interested in just a few scores, but in all the scores in a population. Yet, because behavioral scientists are interested in learning about variables (not constants), it is impossible to know for sure about all the scores in a population from measuring just a few of them. On the other hand, it is time consuming and expensive to collect all the scores in a population, and it may be unethical or impossible. What to do?

samples

The solution to this problem is provided by statistical procedures based on sampling from populations.

> **A sample** is a subset of measurements from a population.

That is, a sample contains some, but usually not all, of the scores in the population. The 608 CESD scores are a sample from the population of CESD scores of all smokers.

An important type of sample is a random sample.

> **A random sample** is selected so that every score in the population has an equal chance of being included.

Whether a sample is random or not does not depend on the actual scores included in the sample, but on how the scores in the sample are selected. Only if the scores are selected in such a way that each score in the population has an equal chance of being included in the sample is the sample a random sample. The CESD scores are not a random sample of CESD scores of all smokers. These scores are only from people living in Madison and Milwaukee, Wisconsin, and there was no attempt to ensure that CESD scores of people living elsewhere were included. Procedures for producing random samples are discussed in Chapter 5.

As you will see in Chapters 5–22, random samples are used to help solve the problem of large populations. That is, with the data in a random sample, we can learn about the population from which the sample was obtained by using inferential statistical procedures.

descriptive and inferential statistical Procedures

descriptive statistical Procedures

Because of variability, in order to learn anything from data, the data must be organized.

> **descriptive statistical procedures** are used to organize and summarize the measurements in samples and populations.

In other words, descriptive statistical procedures do what the name implies—they describe the data. These procedures can be applied to samples and to populations. Most often, they are applied to samples, because it is rare to have *all* the scores in a population.

Descriptive statistical procedures include ways of ordering and grouping data into distributions (discussed in Chapter 2) and ways of calculating single numbers that summarize the whole set of scores in the sample or population (discussed in Chapters 2 and 3). Some descriptive statistical procedures are used to represent data graphically, because as everyone knows, a picture is worth a thousand words.

Inferential Statistical Procedures

The most powerful tools available to the statistician are inferential statistical procedures.

> **Inferential statistical procedures** are used to make educated guesses (inferences) about populations based on random samples from the populations.

These educated guesses are the best way to learn about a population short of collecting all of the scores in the population.

All of this may sound a bit like magic. How can you possibly learn about a whole population that may contain millions and millions (or, theoretically, an infinity) of scores by examining a small number of scores contained in a random sample from that population? It is not magic, however, and it is even understandable. Part II of this book presents a detailed description of how inferential statistical procedures work.

Inferential statistical procedures are so pervasive in our society that you have undoubtedly read about them and made decisions based on them. For example, think about the last time you heard the results of an opinion poll, such as the percentages of the registered voters who favor Candidates A, B, or C. Supposedly, your opinion is included in those percentages (assuming that you are a registered voter so that your opinion is included in the population). But on what grounds does the pollster presume to know your opinion? It is a safe bet that only rarely, if ever, has a pollster actually contacted you and asked you your opinion. Instead, the percentages reported in the poll are educated guesses based on inferential statistical procedures applied to a random sample.

In recent years, it has become fashionable for the broadcast and print media to acknowledge that conclusions from opinion polls are educated guesses (rather than certainties). This acknowledgment is in the form of a "margin of error." The "margin of error" is how much the reported percentages may differ from the actual percentages in the population (see Chapter 11 for details).

Another example of the impact of inferential statistical procedures on our daily lives is in our choices of foods and medicines. Many new food additives and medicines are tested for safety and approved by government agencies such as the Food and Drug Administration (FDA). But how does the FDA know that the new product is safe for you? In fact, the FDA does not know for sure. The decision that a new drug is safe is based on inferential statistical procedures. The FDA example raises several sobering issues about the data used by government agencies to set standards on which our lives literally depend. It is only recently that government agencies have insisted that data be collected from women, and without such data, it is uncertain if a particular drug is actually safe or effective for women. The terrible birth defects attributed to the drug Thalidomide occurred because no one had bothered to collect the data that would verify the safety of the drug with pregnant woman. Similarly, very little data on safe levels of environmental pollutants such as PCBs and pesticides have been collected from children. Consequently, our society may be setting the scene for a disaster by allowing into the environment chemicals that are relatively safe for adults but disastrous for children whose immune systems are immature and whose rapidly developing brains are sensitive to disruption by chemicals.[1]

[1] For an excellent discussion of these issues, see C. F. Moore (2003), *Silent scourge*. New York: Oxford University Press.

The final example of the use of inferential procedures is the behavioral sciences themselves. Most knowledge in the behavioral sciences is derived from data. The data are analyzed using inferential statistical procedures, because interest is not confined to just the sample of scores, but extends to the whole population of scores from which the sample was selected. If you are to understand the data of the behavioral sciences, then you need to understand how statistical procedures work.

Measurement

Data are collected by measuring a variable. But what does it mean to measure a variable?

> **Measurement** is the use of a rule to assign a number to a specific observation of a variable.

As an example, think about measuring the length of your desk. The rule for measuring length is, "Assign a number equal to the number of lengths of a standard ruler that fit exactly from one end of the desk to the other." In this example, the variable being measured is "length." The observation is the length of a specific desk, your desk. The rule is to assign a value (for example, 4 feet) equal to the number of lengths of a standard ruler that fit from one end of the desk to the other.

As another example, consider measuring the weight of a newborn baby. The variable being measured is weight. The specific observation is the weight of the specific baby. The measurement rule is something like, "Put the baby on one side of a balance scale, and assign to that baby a weight equal to the number of pound weights placed on the other side of the scale to get the scale to balance."

Measuring variables in the behavioral sciences also requires that we use a rule to assign numbers to observations of a variable. For example, one way to measure depression is to assign a score equal to the sum of the ratings of the CESD questions. The variable is depression, the specific observation is the depression of the person being assessed, and the rule is to assign a value equal to the sum of the ratings. Similarly, measuring intelligence means assigning a number based on the number of questions answered correctly on an intelligence test.

Considering Measurement in a s ocial and Political Context

The choice of what variables to measure in a study is no accident; usually those choices entail a lot of discussion and planning, and are often influenced by social or political motives of the researcher. The measurement rules, as well, usually involve much discussion, but the details are rarely stated in a study's results. At the very least, there's usually some ambiguity. Take, for example, the LONG variable in the Smoking Study, which measures the longest time without smoking. Let's say that a study participant answers "8 months," which would result in a score of 7 (6–12 months). But, if we probe further,

we may find that the participant actually answered: "Well, I didn't smoke for 4 months, but then one night I had one cigarette, and then didn't have another for 4 months. I say 8 months because it was just a minor slip-up." Is the longest time without smoking for this individual 8 months or 4 months? Is "smoking" defined as "one cigarette" or "one drag" or "buying a pack"? If the researcher is interested in the effectiveness of a particular antismoking program, she may give this participant "a break" and count it as 8 months, because clearly, to her, this participant didn't relapse (it was only one cigarette, after all). A different researcher, interested in showing that all addicts wind up using again (relapsing) might say that one cigarette constitutes a relapse, and score this as 4 months. Political motives may enter a study in this way because for some people the only solution for drug addiction may be abstinence (for example, Alcoholics Anonymous), but for others, recreational drug use may be seen as OK in certain situations (for example, "harm reduction" approaches). In addition, a researcher's grant funding may be dependent on *having* and *solving* a social problem, and maybe even a "growing problem," even though the "problem" is not as big as one might think. Therefore, we should remain critical of how psychologists measure and contemplate what might have been included and what might have been left out.

d ifferences Among Measurement r ules

All rules for measuring variables are not equally good. They differ in three important ways. First, they differ in validity.

> **Validity** refers to how well the measurement rule actually measures the variable under consideration as opposed to some other variable.

Some intelligence tests are better than others because they measure intelligence rather than (accidentally) being influenced by creativity or memory for trivia. Similarly, some measures of depression are better than others because they measure depression rather than introversion or aggressiveness.

Measurement rules also differ in reliability.

> **r eliability** is an index of how consistently the rule assigns the same number to the same observation.

For example, an intelligence test is reliable if it tends to assign the same number to individuals each time they take the test. Books on psychological testing discuss validity and reliability in detail.[2]

Finally, a third difference among measurement rules is that the properties of the numbers assigned as measurements depend on the rule. At first blush, this statement may sound like nonsense. After all, numbers are numbers; how can their properties differ?

[2] A classic text is A. Anastasi (1988), *Psychological testing* (6th ed.). New York: Macmillan.

Properties of Numbers u sed as Measurements

When numbers are measurements, they can have four properties. The first of these is the category property.

> The **category property** is that observations assigned the same number are in the same category, and observations assigned different numbers are in different categories.

For example, suppose that you are collecting data on the types of cars that American citizens drive, and you are most interested in the country in which the cars were manufactured. You could "measure" the country of manufacture (the variable) by using the following rule to assign numbers to observations: If the car was manufactured in the United States, assign it a 1; if manufactured in Japan, assign it a 2; if in Germany, a 3; if in France, a 4; if in Italy, a 5; and if manufactured anywhere else, a 0. These numbers have the category property because each observation assigned the same number (for example, 2) is in the same category (made in Japan).

These country-of-manufacture numbers are different from the numbers that we usually encounter. Typically, assigning a number to an observation (say, Observation A) means more than just assigning observation A to a specific category. For example, if Observation A is assigned a value of 1 and Observation B is assigned a value of 2, it usually means that Observation A is shorter, lighter, or less valuable than Observation B. This is not the case for the measurements of country of manufacture. A car manufactured in the United States (and assigned a number 1) is not necessarily shorter, lighter, or less valuable than a car manufactured in Japan (assigned the number 2). The point is, how we interpret the measurements depends on the properties of the numbers, which in turn depend on the rule used in assigning the numbers.

> Measurements have the **ordinal property** when the numbers can be used to order the observations from those that have the least of the variable being measured to those that have the most.

Consider another example. Suppose that a social psychologist investigating cooperation has a preschool teacher rank the four pupils in the class from least cooperative (first) to most cooperative (fourth). These cooperation scores (ranks) have two properties. First, the scores have the category property, because children assigned different scores are in different categories of cooperation. Second, the scores have the ordinal property because the scores can be used to order the observations from those that have the least to those that have the most cooperation. It is only when measurements have the ordinal property that we know that observations with larger measurements have *more* of whatever is being measured.

A third property that measurements may have is the equal intervals property.

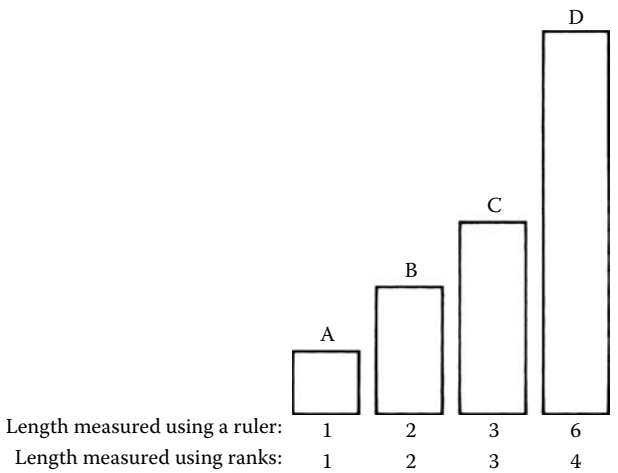
> The **equal intervals property** means that whenever two observations are assigned measurements that differ by exactly one unit, there is always an equal interval (difference) between the observations in the actual variable being measured.

To understand what is meant by equal intervals, consider again measuring the cooperativeness of the four preschool children. The four children (call them Alana, Bob, Carol, and Dan) have cooperation scores of 1, 2, 3, and 4. The difference between Alana's cooperation *score* (1) and Bob's cooperation *score* (2) is 1. Likewise, the difference between Carol's cooperation *score* (3) and Dan's cooperation *score* (4) is 1. The important question is whether the actual difference in *cooperation* (not just the score) between Alana and Bob equals the actual difference in cooperation between Carol and Dan.

It is very unlikely that the difference in cooperation between Alana and Bob equals the difference in cooperation between Carol and Dan. The teacher simply ranked the children from least to most cooperative. The teacher did not take any precautions to ensure equal intervals. Alana and Bob may both be very uncooperative, with Bob being just a bit more cooperative than Alana (the actual difference in cooperation between Alana and Bob is a "bit"). Carol may also be on the uncooperative side, but just a bit more cooperative than Bob (the actual difference between Carol and Bob is a "bit"). Suppose, however, that Dan is the teacher's helper and is very cooperative. In this case, the difference in cooperation between Carol and Dan may be very large, much larger than the difference in cooperation between Alana and Bob. Because the differences in scores are equal (the difference in cooperation *scores* between Alana and Bob equals the difference in cooperation *scores* between Carol and Dan), but the differences in amount of cooperation (the variable) are not equal, these cooperation *scores do not* have the equal interval property.

Now consider using a ruler to measure the lengths of the four lines in Figure 1.1. The lines A, B, C, and D have lengths of 1, 2, 3, and 6 centimeters, respectively. Using a ruler to measure length generates measurements with the equal intervals property: For each pair of observations for which the measurements differ by exactly one unit, the differences in length are exactly equal. That is, the measurements assigned lines A (1) and B (2) differ by one, as do the measurements assigned lines B (2) and C (3); and, important to note, the actual difference in lengths between lines A and B exactly equals the actual difference in length between lines B and C.

FiGu r e 1.1
Length measured using two different measurement rules.



A difficulty in understanding the equal intervals property is in maintaining the distinction between the variable being measured (length or cooperation) and the number assigned as a measurement of the variable. The numbers *represent* or stand for certain properties of the variable. The numbers are not the variable itself. The number 1 is no more the cooperation of Alana (it is a measure of her cooperation) than is the number 1 the actual length of line A (it is a measure of its length). Whether or not the measurements have properties such as equal intervals depends on how the numbers are assigned to represent the variable being measured. Using a ruler to measure length of a desk assigns numbers that have the equal intervals property; using rankings to measure cooperation of preschool children assigns numbers that do not have the equal intervals property.

The difference between the length and cooperation examples is not in what is being measured, but in the rule used to do the measuring. A ranking rule can be used to measure the lengths of lines (this is what we do when we need a rough measure of length—compare two lengths to see which is longer). In this case, the measured lengths of lines A, B, C, and D would be 1, 2, 3, and 4, respectively (see Figure 1.1). These measurements of length do *not* have the equal intervals property, because for each pair of observations for which the measurements differ by exactly one unit, the real differences in length are *not* exactly equal.

The fourth property that measurements may have is the absolute zero property.

> The **absolute zero property** means that a value of zero is assigned as a measurement only when there is nothing at all of the variable that is being measured.

When length is measured using a ruler (rather than ranks), the score of zero is an absolute zero. That is, the value of zero is assigned only when there is no length. When measuring country of car manufacture, zero is not an absolute zero. In that example, zero does not mean that there is no country of manufacture, only that the country is not the United States, Japan, Germany, France, or Italy.

Another example of a measurement scale that does not have an absolute zero is the Fahrenheit (or Centigrade) scale for measuring temperature. A temperature of 0°F does not mean that there is no heat. In fact, there is still some heat at temperatures of –10°F, –20°F, and so on. Because there is still some heat (the variable being measured) when zero is assigned as the measurement, the zero is not an absolute zero.[3]

t ypes of Measurement s cales

In addition to the four properties of measurements (category, ordinal, equal intervals, and absolute zero), there are four types of measurement rules (or scales), determined by the properties of the numbers assigned by the measurement rules.

> **A nominal scale** is formed when the numbers assigned by the measurement rule have only the category property.

[3] The Kelvin scale of temperature does have an absolute zero. On this scale, 0 means absolutely no heat. Zero degrees Kelvin equals –459.69°F.

"Nominal" comes from the word *name*. The numbers assigned using a nominal scale name the category to which the observation belongs but indicate nothing else. Thus, the measurements of country of manufacture of cars form a nominal scale, because the numbers name the category (country), but have no other properties.

Several of the variables in the Smoking Study are measured using nominal scales. For example, TYPCIG (type of cigarette smoked) is measured using a nominal scale defined as 1 = regular filter; 2 = regular no filter; 3 = light; 4 = ultra light; 5 = other. Another nominally measured variable is SPOUSE, that is, whether the smoker's spouse smokes (1) or does not smoke (0). The GENDER variable in the Maternity Study (is the child male or female) is also measured using a nominal scale.

> **An ordinal scale** is formed when the measurement rule assigns numbers that have the category and the ordinal properties, but no other properties.

Many of the variables in the Smoking and Maternity studies are measured using ordinal scales. The longest time without smoking (LONG) variable is measured as 1 = less than a day; 2 = 1–7 days; 3 = 8–14 days; 4 = 15 days to a month; 5 = 1–3 months; 6 = 3–6 months; 7 = 6–12 months; 8 = more than a year. As the assigned score increases from 1 to 8, the length of time without smoking increases, so the numbers have the ordinal property. However, the difference between a measurement of 1 and 2 (LONG 1 – LONG 2 = about 3 days) is not comparable to a difference between a measurement of, say, 5 and 6 (LONG 5 – LONG 6 = about 3 months), thus the measurements do not have the equal intervals property. The researchers might have attempted to measure LONG using a ratio scale by asking participants to estimate the longest number of days without smoking, from 0 to thousands of days. Unfortunately, people's estimates are often clouded by faulty memory processes and faulty estimates. One person who knows that he quit once for more than a year might estimate LONG as 500 days. Another person who had been abstinent for the same amount of time, but who can't remember whether he quit in the year 2001 or 1999, and who can't quite remember how to translate years into days, might estimate LONG as 10,000 days. Thus, these measurements are not as valid or reliable as the simpler ordinal measurements of the LONG scale.

Many behavioral scientists (and businesses that conduct marketing research) collect data by having people rate observations for specific qualities. For example, a clinical psychologist may be asked to rate the severity of his patients' psychopathologies on a scale from 1 (extremely mild) to 10 (extremely severe). As another example, a consumer may be asked to rate the taste of a new ice cream from 1 (awful) to 100 (sublime). In both cases, the measurements represent ordinal properties. For the clinical psychologist, the larger numbers represent more severe psychopathology than the smaller numbers; for the ice-cream raters, the larger numbers represent better-tasting ice cream than the smaller numbers. In neither example, however, do the measurements have the equal intervals property. *As a general rule, ratings and rankings form ordinal scales.*

The third type of scale is the interval scale.

> **interval scales** are formed when the numbers assigned as measurements have the category, ordinal, and equal intervals properties, but not an absolute zero.

Two examples of interval scales are the Fahrenheit and Centigrade scales of temperature. Neither has an absolute zero because 0° (F or C) does not mean absolutely no heat. The measurements do have the category property (all observations assigned the same number of degrees have the same amount of heat), the ordinal property (larger numbers indicate more heat), and the equal intervals property (on a particular scale, a difference of 1° always corresponds to a specific amount of heat).

Many psychological variables are measured using scales that are between ordinal and interval scales. This statement holds for many of the variables included in the Maternity Study, such as marital satisfaction (for example, M1MARSAT), mother's positive affect during free play (MPOS), infant dysregulation during free play (IDYS), and child's internalizing behavior during free play (M7INT). Consider M7INT in a little more detail. To measure the variable, a mother was asked to rate her child's behavior in regard to nine questions such as, "Tends to be fearful or afraid of new things or new situations." The rating scale was 0 = does not apply; 1 = sometimes applies; 2 = frequently applies. Thus, the rating of each question forms an ordinal scale without the equal intervals property. But what happens when we sum the ratings from nine questions to get the M7INT score? It is unlikely that the difference in internalizing behavior between M7INT 10 and M7INT 11 is exactly the same as the difference between, say, M7INT 20 and M7INT 21. Nonetheless, it may well be that these two differences in internalizing behavior are fairly comparable, that is, that the scale is close to having the equal intervals property.

The conservative (and always correct) approach to these "in between" scales is to treat them as ordinal scales. As we will see in Part II, however, ordinal scales are at a disadvantage compared to interval scales when it comes to the range and power of statistical techniques that can be applied to the data. Recognizing this disadvantage, many psychologists treat the data from these in-between scales as interval data; that is, they treat the data as if the measurements were collected using an interval scale. One rule of thumb is that scores from the middle of an in-between scale are more likely to have the equal intervals property than scores from either end. If the data include scores from the ends of an in-between scale, it is best to treat the data conservatively as ordinal.

Many scales for measuring physical qualities (length, weight, time) are ratio scales.

> **A ratio scale** is formed when the numbers assigned by the measurement rule have all four properties: category, ordinal, equal interval, and absolute zero.

The reason for the name "ratio" is that statements about ratios of measurements are meaningful only on a ratio scale. It makes sense to say that a line that is 2.5 centimeters long is *half* (a ratio) the length of a 5-centimeter line. Similarly, it makes sense to say that 20 seconds is *twice* (a ratio) the duration of 10 seconds.

On the other hand, it does not make sense to say that 68°F is twice as hot as 34°F. This is easily demonstrated by converting to Centigrade measurements. Suppose that the temperature of Object A is 34°F (corresponding to 1°C) and that the temperature of Object B is 68°F (corresponding to 20°C). Comparing the amount of heat in the objects using the Fahrenheit measurements seems to indicate that Object B is twice as hot as Object A, because 68 is twice 34. Comparing the measurements on the Centigrade scale (which of course does not change the real amount of heat in the objects), it seems that Object B is 20 times as hot as Object A. Object B cannot be 20 times as hot as object A and at the same time be twice as hot. The problem is that statements about ratios are not meaningful unless

the measurements are made using a ratio scale. Neither ratio (2:1 or 20:1) is right, because neither set of measurements was made using a ratio scale.

This problem does not occur when using a ratio scale. A 5-centimeter (2-inch) line *is* twice as long as a 2.5-centimeter (1-inch) line, and that is true whether the measurements are made in centimeters, inches, or any other ratio measurement of length.

Several variables in the Smoking Study are measured using ratio scales. One example is the carbon monoxide level at the end of treatment measured in parts per million (CO_EOT), and another is the number of times the participant has tried to quit smoking (QUIT).

importance of scale types

The question that may be uppermost in your mind is, "So what?" There are three reasons why knowing about scale types is important. First, now that you know about scale types you will be less likely to make unsupportable statements about data. One such statement is the use of ratio comparisons when the data are not measured using a ratio scale. For example, consider a teacher who gives a spelling test and observes that Alice spelled 10 words correctly, whereas Bill spelled only 5 words correctly. Certainly, Alice spelled twice as many words correctly as did Bill. Nevertheless, the number of words correct on a spelling test is not a ratio measurement of spelling ability (zero words correct does not necessarily mean zero spelling ability). So, although it is perfectly correct to say that Alice spelled twice as many words correctly as did Bill, it is silly to say that Alice is twice as good a speller as is Bill. Similarly, it is not legitimate to claim that a child with an internalizing score (M7INT) of 20 internalizes twice as much as a child with a score of 10.

Second, the types of descriptive statistical procedures that can be applied to data depend in part on the scale type. Although some types of descriptions can be applied to data regardless of the scale type, others are appropriate only for interval or ratio scales, and still others are appropriate for ordinal, interval, and ratio scales, but not nominal scales.

Third, the types of inferential statistical procedures that can be applied to data depend in part on the measurement scale.

Given these three reasons, it is clear that if you want to learn from data you must be able to determine what sort of scale was used in collecting the data. The only way to know the scale type is to determine the properties of the numbers assigned using that rule. If the only property of the measurements is the category property, then the data are nominal; if the measurements have both the category and ordinal properties, then the data are ordinal; if, in addition, the data have the equal interval property, then the data are interval. Only if the data have all four properties are they ratio.

Now that you understand the importance of scale types, it may be helpful to read this section again. Your ability to distinguish among scale types will be used throughout this textbook and in all of your dealings with behavioral data.

using Computers to Learn From data

Data analysis often involves some pretty tedious computations, such as adding columns of numbers. Much of this drudgery can be eliminated by using a computer program such

as Excel, and *Learning From Data* is written to be used with that program. The CD that comes with this book provides the files that your Excel program requires to mesh with the book. First, open up the Read Me file and follow the instructions for loading the Excel Add-ins. These Add-ins provide computer routines that exactly match those used in the book. Second, if you are not familiar with basic Excel operations (e.g, for entering data in a spreadsheet or for selecting rows and columns), you should run the Excel tutorial. Third, the CD includes numerous data files. Two large data files provide the data from the Maternity and Smoking studies. Other data files provide the data used in all of the major worked-out examples and the end-of-chapter exercises.

What Statistical Analysis Programs Can Do for You

The programs have two main benefits. First, they eliminate the drudgery of doing lots of calculations. Second, they ensure accuracy of calculation. A benefit that flows from these two is that the programs make it easy to explore data by conducting multiple analyses.

What the Programs Cannot Do for You

Almost everything that is important is *not* done by the programs. The essence of statistical analysis is choice (choosing the right statistical method and interpretation of the outcome of the chosen method). The programs cannot choose the appropriate methods for you. Similarly, the programs do not know whether a data set is a sample, a random sample, or a population. Consequently, the program cannot adequately interpret the output. *Learning From Data* teaches you how to make good choices and how to interpret the outcome of the statistical methods; the computer eliminates the drudgery.

Because the computer program does the calculations, you might think that you can ignore the formulas in the text. That would be a big mistake for several reasons. First, for small sets of data it is easier to do calculations by hand (or using a calculator) rather than using a computer. But to do the calculations by hand, you need to know the formulas. Second, following the formulas is often the best way to figure out exactly what the statistical technique is doing and how it works. Working through the formulas can be hard intellectual labor, but that is the only way to understand what they do.

SUMMARY

The behavioral sciences are built on a foundation of data. Unfortunately, because behavioral data consist of measurements of variables, individual measurements will differ from one another so that no clear picture is immediately evident. Fortunately, we can learn from variable data by applying statistical procedures.

Descriptive statistical procedures organize, describe, and summarize data. Descriptive statistical procedures can be applied to samples or to populations, but because we rarely have all the scores in a population, descriptive procedures are generally applied to data

from samples. We use inferential statistical procedures to make educated guesses (inferences) about a population of scores based on a random sample of scores from the population. Although these inferences are not error-free, appropriate use of inferential statistical procedures can reduce the chance of error to acceptable levels (for example, the margin of error in a poll).

Appropriateness of a statistical procedure depends in part on the type of measurement scale used in collecting the data. The measurement scale is determined by the properties of the numbers (assigned by the measurement). If the measurements have the category, ordinal, equal interval, and absolute zero properties, then a ratio scale is formed; if the measurements have all but the absolute zero property, an interval scale is formed. If the measurements have only the category and ordinal properties, they form an ordinal scale. Finally, if the measurements have only the category property, they form a nominal scale.

exercises

terms *Define these new terms.*

variable
constant
sample
random sample
population
descriptive statistical procedure
inferential statistical procedure
validity
reliability
measurement
category property
ordinal property
equal intervals property
absolute zero property
nominal scale
ordinal scale
interval scale
ratio scale

Questions *Answer the following questions.*

(Answers are given in the back of the book for questions marked with "†".)

1. Why would there be no need for descriptive or inferential statistical procedures if behavioral scientists could measure constants instead of variables?

2. List 10 different variables and 1 constant in the behavioral sciences.

3. Classify each of the following as a population, a sample, or both. When the answer is both, describe the circumstances under which the data should be considered a population and under which they should be considered a sample.

a. Family incomes of all families in the United States.

†**b.** Family incomes of all families in Wisconsin.

c. The number of words recalled from a list of 50 words by 25 first-year college students who volunteer to take part in an experiment.

d. The number of days spent in intensive care for all people who have undergone heart transplant surgery.

e. The number of errors made by rats learning a maze.

4. Describe two examples of each of the four types of measurement scales. Indicate why each is an example of its type.

†**5.** If you had a choice between using nominal, ordinal, interval, or ratio scales to measure a variable, what would be the best choice? Why?

6. A set of scores can be one type of scale or another, depending on what the set of scores represents. Consider the number of errors made by rats in learning a maze. If the data represent simply the number of errors, then the scores form a ratio scale. The numbers have all four properties, and it makes perfectly good sense to say that if Rat A made 30 errors and Rat B made 15 errors, then Rat A made twice as many errors as Rat B. Suppose, however, that the scores are used as a measure of rat intelligence. Are these scores a ratio measure of intelligence? Explain your answer. What are some of the implications of your answer?

7. Determine the type of measurement scale used in each of the following situations:

a. A supervisor ranks his employees from least to most productive.

†**b.** Students rate their statistics teacher's teaching ability using a scale of 1 (awful) to 10 (magnificent).

c. A sociologist classifies sexual preference as 0 (heterosexual), 1 (homosexual), 2 (bisexual), 3 (asexual), 4 (other).

d. A psychologist measures the time to complete a problem-solving task.

part I

Descriptive Statistics

The three chapters in Part I provide an introduction to descriptive statistical techniques. All of these techniques are designed to help you organize and summarize your data without introducing distortions. As you will see, once the data have been organized, it is far easier to make sense of them; that is, it is far easier to understand what the data are telling you about the world.

Three general types of descriptive techniques are covered. We begin in Chapter 2 with frequency distributions—a technique for arranging the scores in a sample or a population to reveal general trends. We will also learn how to use graphs to illustrate frequency distributions.

A second descriptive technique is computing statistics that summarize frequency distributions with just a few numbers. In Chapter 3, we will learn how to compute several indices of central tendency, the most typical scores in a distribution. We will also learn how to summarize the variability of the scores in a distribution.

Finally, we will consider two methods for describing relative location of individual scores within a distribution—that is, where a particular score stands relative to the others. Percentiles are introduced in Chapter 2. They are often used when reporting the results of standardized tests such as the Scholastic Aptitude Test (SAT) and American College Test (ACT). The other measure of relative standing is the standard score (or z score) discussed in Chapter 4. Standard scores are generally more useful than percentiles, but they require the same background to understand.

All of these descriptive techniques form the underpinning for the remainder of this book, which deals with inferential statistical techniques. Statistical inference begins with a description of the data in a sample, and it is this description that is used to make inferences about a broader population.

Chapter 2

Frequency Distributions and Percentiles

Collecting data means measuring observations of a variable. And, of course, these measurements will differ from one another. Given this variability, it is often difficult to make any sense of the data until they are analyzed and described. This chapter examines a basic technique for dealing with variability and describing data: forming a frequency distribution. When formed correctly, frequency distributions achieve the goals of all descriptive statistical techniques: They organize and summarize the data without distorting the information the data provide about the world.

This chapter also introduces two related topics, graphical representation of distributions and percentiles. Graphical representations highlight the major features of distributions to facilitate learning from the data. Percentiles are a technique for determining the relative standing of individual measurements within a distribution.

While reading this chapter, keep in mind that the procedures for constructing frequency distributions can be applied to populations and to samples. Because it is so rare to actually have all the scores in a population, however, frequency distributions are usually

constructed from samples. Reflecting this fact, most of the examples in the chapter will involve samples.

Frequency distributions

Suppose that you are working on a study of social development. Of particular interest is the age at which aggressive tendencies first appear in children. You begin data collection (measuring the aggressiveness variable) by asking the teacher of a preschool class to rate the aggressiveness of the 20 children in the class using the scale:

Meaning	Score Value
potential for violence	5
very aggressive	4
somewhat aggressive	3
average	2
timid	1
very timid	0

The data are in Table 2.1. As is obvious, the data are variable; that is, the measurements differ from one another. It is also obvious that it is difficult to learn anything from these data as they are presented in Table 2.1. So as a first step in learning from the data, they can be organized and summarized by arranging them in the form of a frequency distribution.

> A **frequency distribution** is a tabulation of the number of occurrences of each score value.

The frequency distribution for the aggressiveness data is given in Table 2.2. The second column lists the score values. The third column in Table 2.2 lists the frequency with which each score value appears in the data. Constructing the frequency distribution involves nothing more than counting the number of occurrences of each score value. There is a simple way to check whether the distribution has been properly constructed: The sum of the frequencies in the distribution should equal the number of observations in the sample (or population). As indicated in Table 2.2, the frequencies sum to 20, the number of observations.

Table 2.1
Aggressiveness ratings for 20 Preschoolers

Child	Rating	Child	Rating	Child	Rating	Child	Rating
a	4	f	0	k	3	p	2
b	3	g	3	l	0	q	3
c	1	h	3	m	4	r	3
d	1	i	4	n	2	s	1
e	2	j	2	o	3	t	3

t Ab Le 2.2
Frequency d istributions for the Aggressiveness d ata in t able 2.1

Meaning	Score Values	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
Very Timid	0	2	.10	2	.10
Timid	1	3	.15	5	.25
Average	2	4	.20	9	.45
Aggressive	3	8	.40	17	.85
Very Aggressive	4	3	.15	20	1.00
Potential for Violence	5	0	.00	20	1.00
		20	1.00		

It is clear that the frequency distribution has a number of advantages over the listing of the data in Table 2.1. The frequency distribution organizes and summarizes the data, thereby highlighting the major characteristics. For example, it is now easy to see that the measurements in the sample range from a low of 0 to a high of 4. Also, most of the measurements are in the middle range of score values, and there are fewer measurements in the ends of the distribution.

Another benefit provided by the frequency distribution is that the data are now easily communicated. To describe the data, you need to report only five pairs of numbers (score values and their frequencies).

Try not to confuse the numbers representing the score values and the numbers representing the frequencies of the particular score values. For example, in Table 2.2 the number "4" appears in the column labeled "score value" and the column labeled "frequency." The meaning of this number is quite different in the two columns, however. The score value of 4 means a particular level of aggressiveness (*very aggressive*). The frequency of 4 means the number of times a particular score value was observed in the data. In this case, a score value of 2 (*average*) was observed four times.

To help overcome any confusion, be sure that you understand the distinctions among the following terms. "Score value" refers to a possible value on the measurement scale. Not all score values will necessarily appear in the data, however. If a particular score value is never assigned as a measurement (for example, the score value 5, *potential for violence*), then that score value would have a frequency of zero. "Frequency" refers to the number of times a particular score value occurs in the data. Finally, the terms "measurement," "observation," and "score" are used interchangeably to refer to a particular datum—the number assigned to a particular individual. Thus, in Table 2.2, the score value of 1 (*timid*) occurs with a frequency of 3. Similarly, there are three scores (measurements, observations) with the score value of 1 (*timid*).

r elative Frequency

An important type of frequency distribution is the relative frequency distribution.

> **r elative frequency** of a score value is the proportion of observations in the distribution at that score value. **A relative frequency distribution** is a listing of the relative frequencies of each score value.

The relative frequency of a score value is obtained by dividing the score value's frequency by the total number of observations (measurements) in the distribution. For example, the relative frequency of aggressive children (score value of 3) is 8/20 = .40.

Relative frequency is closely related to percentage. Multiplying the relative frequency by 100 gives the percentage of observations at that score value. For these data, the percentage of children rated aggressive is .40 × 100 = 40%.

The fourth column in Table 2.2 is the relative frequency distribution for the aggressiveness data. Note that all of the relative frequencies are between 0.0 and 1.0, as they must be. Also, the sum of the relative frequencies in the distribution will always equal 1.0. Thus, computing the sum is a quick way to ensure that the relative frequency distribution has been properly constructed.

Relative frequency distributions are often preferred over raw frequency distributions because the relative frequencies combine information about frequency with information about the number of measurements. This combination makes it easier to interpret the data. For example, suppose that an advertisement for Nationwide Beer informs you that in a "scientifically selected" sample, 90 people preferred Nationwide, compared to only 10 who preferred Brand X. You may conclude from these data that most people prefer Nationwide. Suppose, however, that the sample actually included 10,000 people, 90 of whom preferred Nationwide, 10 of whom preferred Brand X, and 9,900 of whom could not tell the difference. In this case, the relative frequencies are much more informative (for the consumer). The relative frequency of preference for Nationwide is only .009.

The same argument in favor of relative frequency can also be made (in a more modest way) for the data on aggressiveness. It is more informative to know that the relative frequency of aggressive children is .15 than to simply know that three children were rated as aggressive.

When describing data from *random* samples, relative frequency has another advantage. The relative frequency of a score value in a random sample is a good guess for the relative frequency of that score value in the population from which the random sample was selected. There is no corresponding relation between frequencies in a sample and frequencies in a population.

Cumulative Frequency

Another type of distribution is the cumulative frequency distribution.

> **A cumulative frequency distribution** is a tabulation of the frequency of all measurements at or smaller than a given score value.

The fifth column in Table 2.2 is the cumulative frequency distribution for the aggressiveness scores. The cumulative frequency of a score value is the frequency of that score value plus the frequency of all smaller score values. The cumulative frequency of a score value of zero (*very timid*) is 2. The cumulative frequency of a score value of 1 (*timid*) is

obtained by adding 3 (the frequency of *timid*) plus 2 (the frequency of *very timid*) to get 5. Note that the cumulative frequency of the largest score value (5) equals 20, the total number of observations. This must be the case, because cumulative frequency is the frequency of all observations at smaller than a given score value, and all of the observations must be at or smaller than the largest score value. Also, note that the cumulative frequencies can never decrease when going from the lowest to the highest score value. The reason is that the cumulative frequency of the next higher score value is always obtained by *adding* to the lower cumulative frequency.

The notion of "at or smaller" implies that the score values can be ordered so that we can determine what is "smaller." Thus, cumulative frequency distributions are usually not appropriate for nominal data.

> **A cumulative relative frequency** distribution is a tabulation of the relative frequencies of all measurements at or below a given score value.

The last column in Table 2.2 lists the cumulative relative frequencies for the aggressiveness data. These numbers are obtained by adding up the relative frequencies of all score values at or smaller than a given score value.

Cumulative frequency distributions are most often used when computing percentiles. We shall postpone further discussion of these distributions until that section of the chapter.

Grouped Frequency Distributions

The aggressiveness data were particularly amenable to description by frequency distributions in part because there were only a few score values. Sometimes, however, the data are not so accommodating, and a more sophisticated approach is called for.

Consider, for example, the first 60 measurements on the YRSMK variable in the Smoking Study (Table 2.3). Because the measurements are variable, it is difficult to learn anything from the data as presented in this table.

The frequency distribution is presented in Table 2.4. As you can see, the frequency distribution for these data does not provide a very useful summary of the data. The problem is that there are too many different score values.

TABLE 2.3
YRSMK—Number of years smoking daily From the First 60 Participants in the Smoking Study

5	13	17	20	19	35	21	28	3	22
26	13	30	30	30	32	40	27	14	4
27	33	28	45	29	25	38	35	33	39
5	4	20	24	25	27	16	25	38	9
36	20	18	11	12	23	22	27	32	49
22	30	0	32	4	23	9	29	22	23

t Ab Le 2.4
Frequency d istribution of First 60 yrs Mk s cores

Score Value	Frequency	Score Value	Frequency	Score Value	Frequency
0	1	17	1	34	0
1	0	18	1	35	2
2	0	19	1	36	1
3	1	20	3	37	0
4	3	21	1	38	2
5	2	22	4	39	1
6	0	23	3	40	1
7	0	24	1	41	0
8	0	25	3	42	0
9	2	26	1	43	0
10	1	27	4	44	0
11	1	28	2	45	1
12	0	29	2	46	0
13	2	30	4	47	0
14	1	31	0	48	0
15	0	32	3	49	1
16	1	33	2		

The solution is to group the data into clusters called class intervals.

> **A class interval** is a range of score values. **A grouped frequency distribution** is a tabulation of the number of measurements in each class interval.

The grouped frequency distribution is presented in Table 2.5. The class intervals are listed on the left. The lowest interval, 0–4, contains all of the measurements between (and including) 0 and 4. The next interval, 5–9, contains the measurements between 5 and 9, and so on.

Clearly, the data in the grouped distribution are much more easily interpreted than when the data are ungrouped. We can now see that most of the people in this sample have been smoking for 20–30 years, although there are a few who have been smoking for more than 45 years and a few who have been smoking only a couple of years.

Relative and cumulative frequency distributions can also be formed from grouped data. Relative frequencies are formed by dividing the frequency in each class interval by the total number of measurements. Cumulative distributions are formed by adding up the frequencies (or relative frequencies) of all class intervals at or below a given class interval. These distributions are also given in Table 2.5.

Constructing Grouped d istributions

A grouped frequency distribution should summarize the data without distorting them. Summarization is accomplished by forming class intervals; if the intervals are inappropriate (for

TABLE 2.5
Grouped Frequency distributions for yrs Mk scores in table 2.3

Class Interval	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
0–4	5	.083	5	.083
5–9	4	.067	9	.150
10–14	5	.083	14	.233
15–19	4	.067	18	.300
20–24	12	.200	30	.500
25–29	12	.200	42	.700
30–34	9	.150	51	.850
35–39	6	.100	57	.950
40–44	1	.017	58	.967
45–49	2	.033	60	1.00
Total	60	1.000		

TABLE 2.6
Grouped Frequency distributions for yrs Mk scores in table 2.3

Interval	*f* (YRSMK)
0–19	18
20–39	39
40–59	3
Total	60

example, too big), however, the data are distorted. As an example of distortion, Table 2.6 summarizes the YRSMK data from Table 2.3 using three large intervals. Indeed, the data are summarized, but important information regarding how the measurements are distributed is lost. The following steps should be used to construct grouped distributions that summarize but do not distort.

Guidelines for grouped frequency intervals:

1. There should be between 8 and 15 intervals.
2. Use convenient class interval sizes, like 2, 3, 5, or multiples of 5.
3. Start the first interval at or below your lowest score.

To construct a good grouped frequency distribution:

1. Compute the range of your scores by subtracting the lowest score from the highest score (Range = High Score – Low Score).
2. Divide the range by 8 and 15. Find a convenient number in between those two values. That will be your class interval. This is also known as your "bin width."

3. Select a starting value. The starting value could be your lowest score, but if your class interval is a multiple of 5, then you may want to select a more convenient, and hence lower starting point. For example, the intervals 0–9, 10–19, 20–29, etc., work very well if you have determined that a class interval of 10 is appropriate. If your lowest score is 3, the intervals 3–12, 13–22, 23–32, etc., do not seem as intuitive as 0–9, 10–19, 20–29, etc. (or 1–10, 11–20, 21–30, etc.).
4. Beginning with your starting value, construct intervals of increasing value.
5. Count the number (frequency) of scores in each interval.

One other step is needed when the measurements contain decimals instead of whole numbers. In these cases, all of the measurements should be rounded so that they have the same number of decimal places.

These steps were used to construct the grouped frequency distribution in Table 2.5. For Step 1, the range was computed as 49 (49 – 0). For Step 2, the range was divided by 8 (49/8 = 6.125) and 15 (49/15 = 3.267), and a convenient number in between those two (5) was selected as the class interval. For Step 3, because the lowest score was 0, the starting value was set at 0. For Step 4, starting with 0, consecutive intervals, of width 5, were constructed: 0–4, 5–9, 10–14, etc.

Note that the interval 5–9 includes the five score values 5, 6, 7, 8, and 9. Thus, the interval size really is 5, even though the difference between 9 and 5 is 4.

Once the lowest interval is specified, the remaining class intervals are easily constructed. Each successive interval is formed by adding the interval size (5) to the bounds of the preceding interval. For example, the interval 10–14 was obtained by adding 5 to both the lower and upper bounds of the interval 5–9. Finally, tabulate the number of measurements within each interval to construct the frequency distribution.

For a second example of grouping, consider the data in Table 2.7. The 60 measurements in this table are from the Smoking Study. Each measurement is a participant's score on the Wisconsin Inventory of Smoking Dependence Motives (WISDM), which are ratings on 65 questions such as "Does smoking make a good mood better?"

TABLE 2.7
First 60 Scores on the Wisconsin Inventory of Smoking Dependence Motives (WISDM)

52.9952	60.7071	53.2262	82.0333	65.9119	59.7071
44.4405	38.3167	62.2333	39.6786	46.2762	52.1119
68.4571	66.4476	28.2667	60.6667	50.0857	44.5690
33.6786	55.1119	21.8190	27.6929	53.1310	36.5500
57.4857	63.9262	60.0548	50.8071	61.2405	66.5810
55.3071	28.4643	43.9143	67.8524	54.7310	52.9429
60.8405	60.7238	51.0786	35.5071	54.2524	65.5429
60.1310	78.9357	65.1976	32.4833	51.2381	48.5786
62.5905	80.6071	54.0476	68.8190	52.1738	55.4214
61.4619	53.3571	35.8976	59.3190	68.1143	62.9429

Because these measurements contain decimals, we begin by making sure that all have the same number of decimal places, as they do. For Step 1, we compute the range of scores by subtracting the lowest score (21.8190) from the highest score (82.0333) to arrive at 60.2143.

In Step 2, we divide the range by 8 and 15: 60.2143/8 = 7.526788 and 60.2143/15 = 4.041287. We choose a number between these two results, preferably a multiple of 2, 3, or 5. We might choose 5.6901, which is a number between 7.526788 and 4.041287, and is divisible by 3, but 5.6901 will not serve as a convenient class interval. Rather, 5 is between 7.526788 and 4.041287, divisible by 5 (obviously), and convenient.

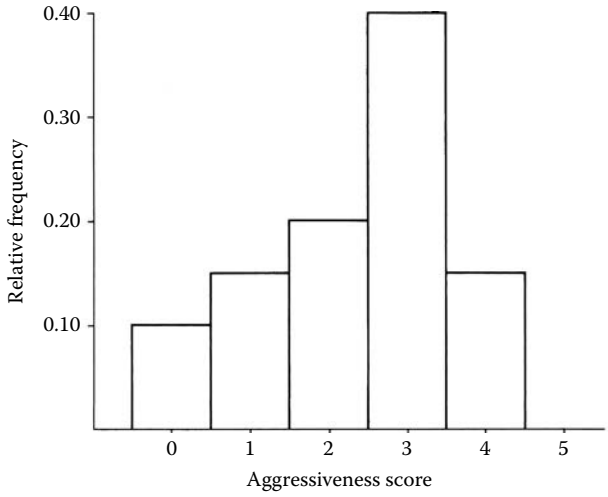
The third step, selecting a starting value, could be set at the lowest score, 21.8190, but 20.0000 seems more intuitive. The first interval, therefore, will be 20.0000–24.9999, the next 25.0000–29.9999, and so on. The final step is to tabulate the number of measurements in each interval to obtain the frequency distribution, and then divide each frequency by the total number of observations to obtain the relative frequency distribution.

As you can tell from Table 2.8, these data are very interesting. The distribution appears "top heavy." In other words, more than half of the scores are greater than 50. This may not be unexpected, though, for it is a measure of "smoking motives" and smokers (which all the participants in the study are) may have many motives to smoke. Nevertheless, these data may be important to the study's designers because they can show that their participants were highly motivated to smoke, as opposed to participants who weren't motivated to smoke. In the end, the study's authors, if the experiment is successful, can claim that their intervention works for people highly motivated to smoke.

TABLE 2.8
Relative Frequency Distribution for WISD Scores (First 60 Subjects)

Class Interval	Relative Frequency
20.0000–24.9999	0.017
25.0000–29.9999	0.050
30.0000–34.9999	0.033
35.0000–39.9999	0.083
40.0000–44.9999	0.050
45.0000–49.9999	0.033
50.0000–54.9999	0.233
55.0000–59.9999	0.100
60.0000–64.9999	0.200
65.0000–69.9999	0.150
70.0000–74.9999	0.000
75.0000–79.9999	0.017
80.0000–84.9999	0.033
Total	1.000

FiGu r e 2.1
r elative frequency histogram for the aggressiveness scores in t able 2.2.



Gr APh iNG Fr e Qu e NCy dist r ibut io Ns

Displaying a frequency distribution as a graph can highlight important features of the data. Graphs of frequency distributions are always drawn using two axes.

> The **abscissa** or **x-axis** is the horizontal axis. For frequency and relative frequency distributions, the abscissa is marked in units of the variable being measured, and it is labeled with the variable's name. The **ordinate** or **y-axis** is marked in units of frequency or relative frequency, and so labeled.

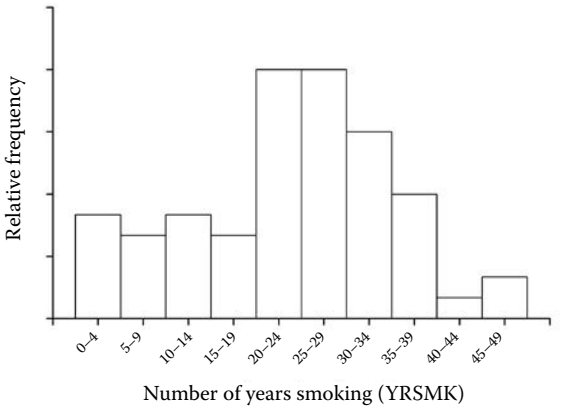
In Figure 2.1, the abscissa is labeled with values of the aggressiveness variable for the distribution in Table 2.2. The ordinate is marked to represent relative frequency of the measurements. Techniques for graphing frequency and relative frequency distributions are almost exactly the same. The only difference is in how the ordinate is marked. Because relative frequency is generally more useful than raw frequency, the examples that follow are for relative frequency distributions.

h istograms

Figure 2.1 is a relative frequency histogram for the aggressiveness data.

> **A relative frequency histogram** uses the heights of bars to represent relative frequencies of score values (or class intervals).

FiGu r e 2.2
r elative frequency histogram for the number of years smoking scores in t able 2.5.



To construct the histogram, place a bar over each score value. The bar extends up to the appropriate frequency mark on the ordinate. Thus, a bar's height is a visual analogue of the score value's relative frequency: the higher the bar, the greater the relative frequency.

Relative frequency histograms can also be drawn for grouped distributions. For these distributions, a bar is placed over each class interval.

Figure 2.2 is a relative frequency histogram of the YRSMK scores in Table 2.5. Sometimes, only the midpoints of each interval are shown on the abscissa. The midpoint of a class interval is the average of the interval's lower bound and the upper bound. Again, the height of each bar corresponds to its relative frequency.

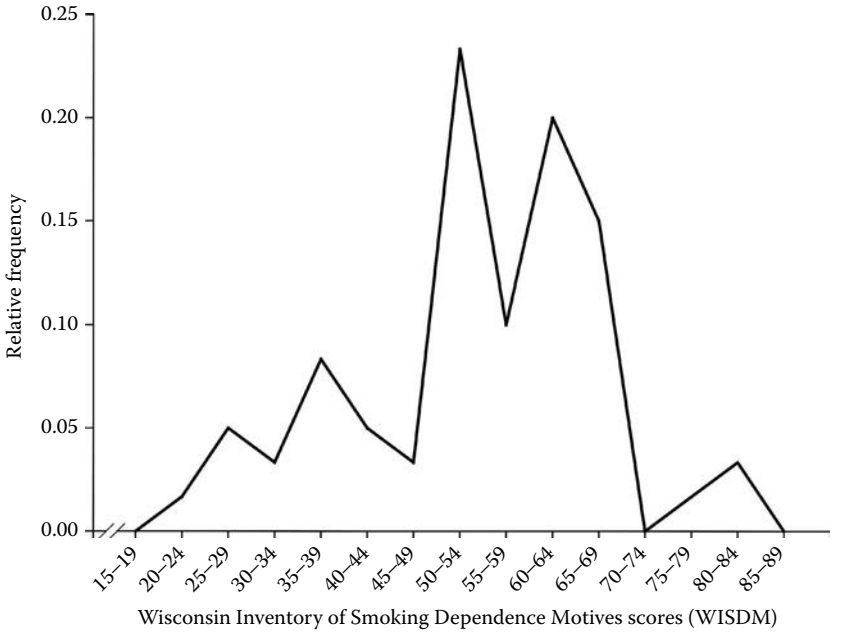
The relative frequency histogram illustrated in Figure 2.2 makes particularly clear some of the salient characteristics of the distribution. For example, it is easy to see that most of the scores are in the middle of the distribution and that there is a decrease in frequency from the moderate scores to the higher scores.

Frequency Polygons

Figure 2.3 is an example of a relative frequency polygon using the WISDM scores in Table 2.8. The axes of a relative frequency polygon are the same as for a histogram. However, instead of placing a bar over each midpoint (or score value), a dot is placed over the midpoint so that the height of the dot corresponds to the relative frequency of the class interval. Next, adjacent dots are connected by straight lines.

As a convention, an additional midpoint (or score value) is added to each end of the distribution (the intervals "15–19" and "85–89" in Figure 2.3). These extra intervals are always drawn with a frequency of zero because no measurements are actually in the intervals. Using the zero-frequency midpoints "closes" the figure made by connecting the dots,

Figure 2.3
Relative frequency polygon for the scores on the WISDM in Table 2.8.



producing a more visually pleasing figure. (A many-sided closed figure is a polygon, which is why this sort of graph is called a relative frequency polygon.)

Traditionally, the abscissa and the ordinate are drawn so that they intersect at a value of 0 on the ordinate and a value of 0 on the abscissa. In Figure 2.3, the double slash marks on the abscissa indicate that there is a "break" in the axis so that the first mark, 15, is not actually 15 units from the intersection.

In general, graphs highlight the salient characteristics of distributions more effectively than do tables. Comparison of Figure 2.3 and the distribution in Table 2.8 demonstrates this point nicely. Starting with the table, it takes some effort to appreciate that the distribution has high frequencies of high scores (intervals 50–54, 60–64, etc.). Figure 2.3 portrays this unusual quality dramatically and without requiring any effort to appreciate it.

When to Use Histograms and Frequency Polygons

We must answer two questions: When should we use graphing techniques? Given that a graph is appropriate, when should a histogram be used and when should a polygon be used? In answer to the first question, graphs of distributions are used whenever it is important to highlight features of the distribution such as the shape, the range (the number of score values), and the location of the distribution on the measurement scale (the typical or middle score values). Each of these features is easily grasped from a picture, but harder to obtain from just the tabular form of the distribution.

Often, graphs are used when two or more distributions must be compared. Because information such as shape of the distribution is easily obtained from a graph, you can actually see that graphs of two distributions are similar or dissimilar in shape. Comparison of two distributions in tabular form is usually more difficult.

The choice between the use of histograms and polygons is often a matter of personal taste. There is one generally accepted rule, however, that depends on the distinction between discrete and continuous variables.

> A **discrete variable** can take on a limited number of score values (such as whole numbers) and can be measured exactly. **A continuous variable** can take on any score value and requires an infinite number of decimal places to specify.

The distinction between continuous and discrete does not depend on the measurement scale, but on the nature of the variable. For example, the variable "family size" is a discrete variable because it can take on only a limited number of score values (whole numbers). This is true whether we measure family size using an ordinal scale (small, medium, or large) or a ratio scale (the actual count of family members).

On the other hand, a variable such as "time to make a grammaticality judgment" is a continuous variable—time can take on an unlimited number of score values. The variable is continuous even if our measuring device (clock) gives measurements only to the nearest whole second.

The rule for deciding between histograms and polygons is that histograms should be used when the variable is discrete or when the variable is measured on a nominal scale; otherwise, it may be preferable to use a polygon. The reason for this is simple. Connecting the points together in the polygon suggests that there are possible score values between the points. Furthermore, the lines connecting the points in a polygon suggest that the relative frequencies of these in-between score values can be estimated by the heights of the lines. Indeed, these suggestions are often appropriate for continuous variables. However, these suggestions are misleading for discrete variables, because there are no score values in between those indicated in the graph. Thus, histograms should be used for discrete variables.

Characteristics of Distributions

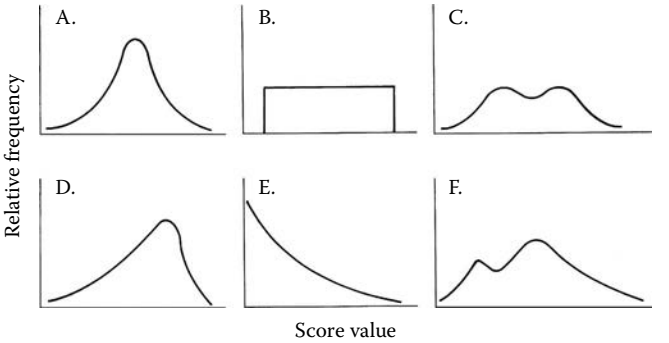
Distributions differ in three major characteristics: shape, central tendency, and variability. Over the next few pages, we will practice comparing distributions by using these characteristics and you will see that they play a major role in the study of statistics.

All of the illustrations that follow will use relative frequency polygons. However, dots will not be placed over specific score values, and continuous lines will be used to illustrate the general shapes of the distributions. The purpose of this departure from standard procedure is to illustrate general principles that do not depend on specific distributions.

Shape

The shape of a distribution can be broadly classified as symmetric, positively skewed, or negatively skewed.

FiGu r e 2.4
r elative frequency distributions of various shapes.



A symmetric distribution can be divided into two halves that are mirror images of each other.

The distributions illustrated in the top row of Figure 2.4 are symmetric distributions. The distribution of years smoking in Figure 2.2 can be characterized as "somewhat symmetric."

In contrast, a skewed distribution cannot be divided into halves that are mirror images. The distributions illustrated in the bottom row of Figure 2.4 are skewed, as is the distribution of WISDM scores in Figure 2.3.

A positively skewed distribution has score values with low frequencies that trail off toward the positive numbers (to the right). **A negatively skewed distribution** has score values with low frequencies that trail off toward the negative numbers (to the left).

Distributions E and F in Figure 2.4 are positively skewed; Distribution D has a negative skew. Note that a skewed distribution has one "tail" that is longer than the other. If the longer tail is pointing toward the positive numbers, then the distribution has a positive skew; if the longer tail is pointing toward the negative numbers, then the distribution has a negative skew.

Salary distributions are often positively skewed. Salaries cannot be less than $0.00, so the tail on the left cannot trail off very far. Most people have salaries in the midrange, but some people have very large salaries, and some (although relatively few) have enormous salaries. The people with enormous salaries produce a positive skew in the distribution.

Modality is another aspect of the shape of a distribution. Modality is the number of clearly distinguishable peaks in the distribution.

A **unimodal distribution** has one peak. A **bimodal distribution** has two peaks. A **multimodal distribution** has more than two peaks.

In Figure 2.4, Distributions A, D, and E are unimodal, and Distributions C and F are bimodal.

Some shapes of distributions have special names. A distribution that is unimodal and symmetric (such as Distribution A in Figure 2.4) is called a bell-shaped distribution. (A normal distribution is a special type of bell-shaped distribution that we will discuss in Chapter 4.) Many psychological variables (such as intelligence) have bell-shaped distributions.

Distribution B in Figure 2.4 is called a rectangular distribution. Note that this distribution is symmetric, but it does not have a well-defined mode. Rectangular distributions indicate that all of the score values have the same relative frequency. Rectangular distributions often arise in gambling situations, such as tossing a fair coin (relative frequency of heads = relative frequency of tails = .5) and rolling a fair die (each number has a relative frequency of one sixth).

Distribution E in Figure 2.4 is called a J-curve (although backward J-curve would be more appropriate). This distribution is positively skewed, and is frequently seen in the relative frequencies of rare events, such as number of lotteries won. Most of us have never won a lottery, so the score value of zero has the greatest relative frequency. Some people have won one or two lotteries, and a few have won even more, producing the long tail on the right.

Central t endency

In addition to shape, distributions differ in central tendency.

> The **central tendency of a distribution** is the score value near the center of the distribution. It is a typical or representative score value.

You may think of the central tendency as the score value that is most representative of the distribution—that is, the score value that you would choose to give the general flavor of the distribution. On the other hand, you may think of the central tendency as the location of the center of the distribution on the measurement scale (represented by the abscissa in a graph). In Chapter 3, we will discuss the mean, the median, and the mode, which are numerical indices of central tendency.

The distributions in the top row of Figure 2.5 have the same shape, but differ in central tendency (the central tendencies are indicated by arrows). In the bottom of the figure, the distributions differ in both shape and central tendency. In both the top and the bottom, the central tendency increases from left to right.

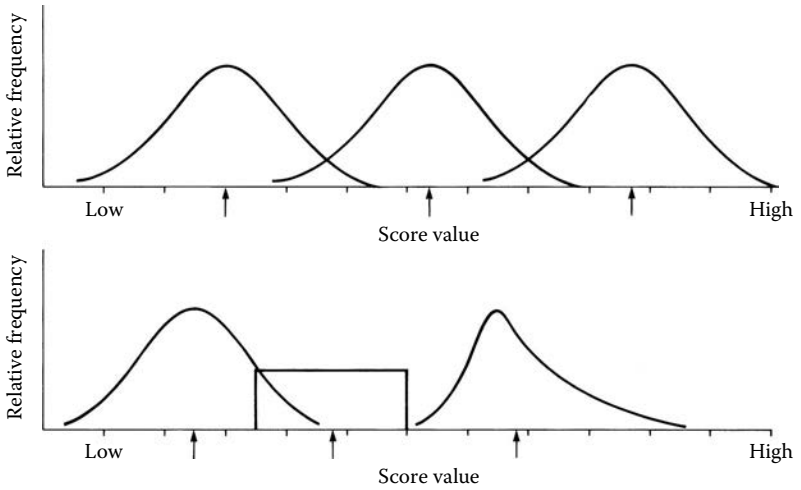
Variability

As you will recall, variability is the driving force behind statistics. Indeed, distributions arise because we measure variables so that not all the measurements are the same.

> **Variability** is the degree to which the measurements in a distribution differ from one another.

When all the measurements in a distribution are similar, the distribution has little variability. In fact, in the rare instance when all of the measurements in the distribution are the

Figure 2.5
Top: **distributions that differ only in central tendency.** ***Bottom:*** **distributions that differ in shape and central tendency.**



same, the distribution has no variability. When the measurements in a distribution deviate greatly from one another, the distribution has much variability. In Chapter 3, we will learn how to compute the variance and the standard deviation, two particularly useful numerical indices of variability.

The top half of Figure 2.6 illustrates how distributions can have the same general shape and the same central tendency, but differ in variability. Variability increases from Distribution A to B to C.

The bottom of Figure 2.6 illustrates how distributions can have different central tendencies and different variabilities. From left to right, the distributions increase in central tendency. However, Distribution D has the greatest variability and Distribution E has the least variability.

Comparing distributions

Now that you have learned how distributions can differ, you have the skills needed to quickly compare and summarize distributions. Simply determine how the distributions compare in terms of shape, central tendency, and variability.

Recall that in the beginning of this chapter we started with an example of a psychologist investigating the development of aggressive behavior. Suppose that the psychologist had segregated the data by sex of the children and had constructed separate relative frequency distributions for the girls and the boys. Figure 2.7 illustrates two possible distributions.

The distribution for the girls is more symmetric than that for the boys; the distribution for the boys is positively skewed. Both distributions are unimodal. The two distributions have similar central tendencies, but they differ in variability; the distribution for the boys is more variable than that for the girls.

FiGure 2.6
t op: **distributions that differ only in variability.** ***Bottom:*** **distributions that differ in both variability and central tendency.**

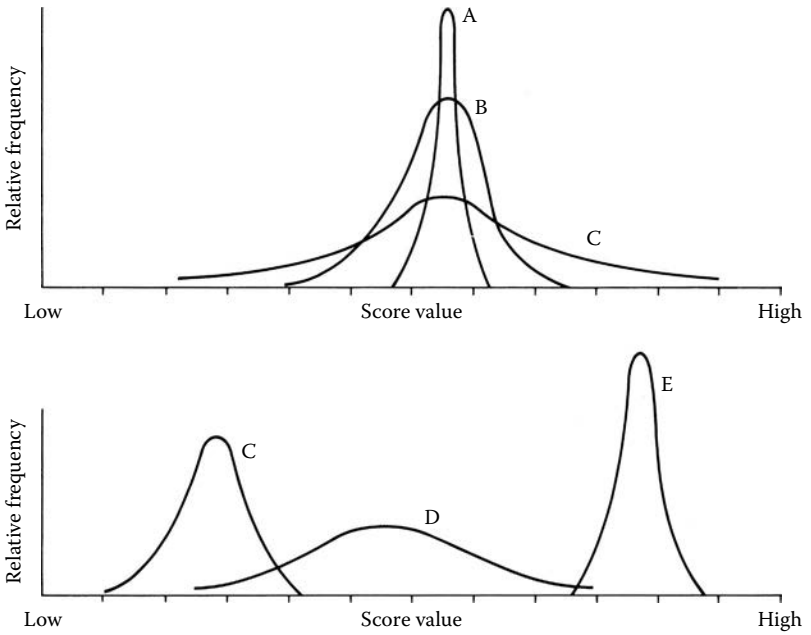
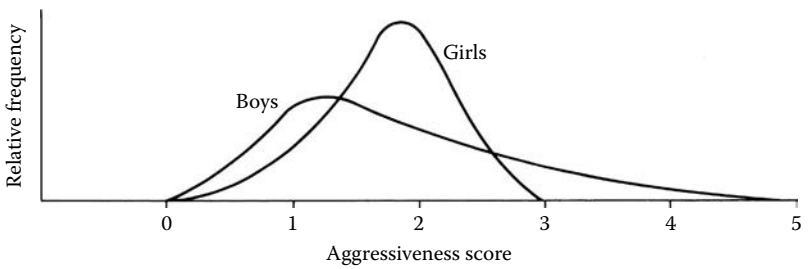


FiGure 2.7
distributions of aggressiveness scores for boys and girls.



Now that the data have been described, it is up to you to determine their importance. Why is the distribution for the boys more variable than that for the girls? Does this represent a basic difference between the sexes, or is it due to socialization? Should these data be used to argue for social change? Of course, these particular questions are not statistical questions, and statistical analyses will not provide answers to them. The point is that statistical analysis can transform the data into a form that helps *you* to ask good questions. But it is up to you to ask the questions.

Percentiles

We have been working with whole distributions and paying little attention to individual measurements within a distribution. At times, however, the individual measurements are of great importance. Consider this example. One of your professors has just handed back your exam. On the top, in bold red ink, is the number 32. At first you are worried. But when you see the 25 on your neighbor's paper, you begin to feel a little better. Still, is 32 a good score?

How good an exam score is depends in part on how you define "good." As a first approximation, you might think that a good score on a test corresponds to achieving close to the maximum possible. Thus, if the exam had a total of 50 points possible, your score of 32 corresponds to 64% correct, which is not very good.

Suppose, however, that the exam was extremely difficult and your score of 32 is one of the best in the class. In this case, you are justified in thinking that your score of 32 (or 64%) is really very good.

The point is this: An informative index of the goodness of a score is the standing of that score relative to the other scores in a distribution. Scores that are near the top of the distribution are "good," regardless of the actual score value or the percent correct. Scores near the bottom of the distribution are "poor," regardless of the actual score value. (If you obtained 90% correct on a test, that would be a "poor" score if everyone else in the class obtained 100% correct.)

Percentile ranks provide just such an index of goodness by giving the *relative standing* of a score—the score's location in the distribution relative to the other measurements in the distribution. More formally:

> The **percentile rank** of a score value is the percent of measurements in the distribution below that score value.

Note that percentile rank and percent correct are not the same. A score value of 80% correct might have any percentile rank between 0 and 100, depending on the number of measurements in the distribution less than 80% correct.

When exam grades are given as percentiles, you have an easily interpreted index of relative standing—how well you did relative to the others in the distribution. If your score has a percentile rank of 95%, then you did better than 95% of the others in the distribution. If your score has a percentile rank of 30%, you did better than 30% of the others.

Percentile ranks are often used to report the results of standardized tests such as the SAT (Scholastic Aptitude Test) and the GRE (Graduate Record Examination). Your percentile rank indicates the percentage of students who received scores lower than yours. So, in terms of your relative performance, you should be happier with higher percentile ranks than lower percentile ranks.

Of course, percentile ranks are not always associated with just exam scores. As we will see, percentile ranks can be determined for scores in any distribution (as long as the measurement scale is not nominal). Thus, you can determine the percentile rank of an aggressiveness score of 4, or the percentile rank of a WISDM score of 62.4963.

Percentile r anks and Percentiles

As you now know, the percentile rank of a score value indicates the percent of measurements smaller than that score value. In the context of percentile ranks, the score values themselves are often called percentiles.

> The *P*th **percentile** is the score value with *P%* of the measurements in the distribution below it.

t hree Precautions

There are three things to be aware of when you use percentiles and percentile ranks. First, percentile ranks and percentiles are only approximate; for grouped distributions, particularly, answers will be approximate.

Second, percentile ranks are an ordinal index of relative standing. That is, the percentile rank of a score value indicates that *P%* of the measurements are smaller than the score value; the percentile rank does not indicate *how* far below.

Third, percentile ranks can be interpreted only within the context of a specific distribution. An example will help to make this precaution clear. Suppose that your score on an English-language achievement test has a percentile rank of 15% (that is, your score is in the 15th percentile). Such a low score might at first be alarming.

However, if all the other measurements in the distribution come from graduate students in English, your score is perfectly reasonable. That same score value may be in the 75th percentile if the test were given to high school students. Remember, percentile ranks are a measure of relative standing—the location of a particular score value relative to the other measurements in the *distribution*. Thus, interpretation of a percentile rank depends on careful consideration of the sample (or population) described by the distribution.

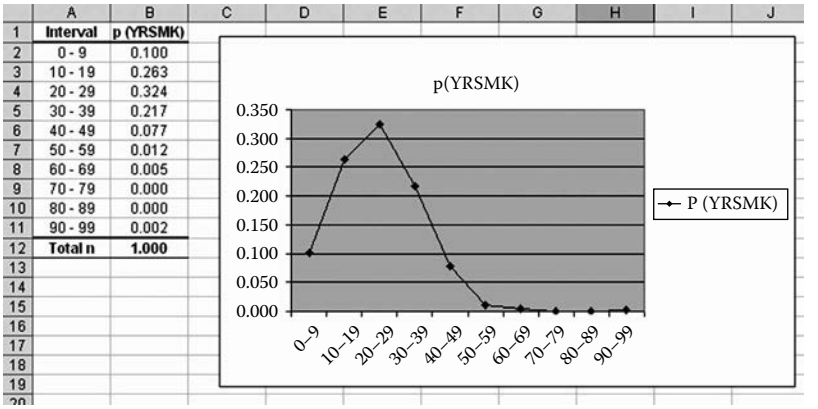
Co MPu t At io Ns u s iNG e x Ce L

Constructing Frequency d istributions

Constructing a frequency distribution can be a tedious enterprise, especially if a data set contains a large number of scores. Take, for example, the Smoking Study data on the CD provided with the book. In a previous example, we looked at the first 60 scores on the YRSMK variable (Number of Years Smoking Daily) from the Smoking Study, but there are 608 total scores in the data set. It would probably take hours to construct a frequency distribution, and then there will likely be errors.

Constructing a frequency distribution in Excel can be accomplished using the LFD3 Analysis Add-in. Open the Smoking data set in Excel, then click on "Tools" and select "LFD Analyses," which will bring up a window with a number of additional analysis tools. "Frequency Distribution" is the first option. When you select this option, a new window

FiGu r e 2.8
Completed frequency distribution analysis using y r s Mk data and the LFd 3 Add-in.



	Interval	p (YRSMK)
1	0 - 9	0.100
2	10 - 19	0.263
3	20 - 29	0.324
4	30 - 39	0.217
5	40 - 49	0.077
6	50 - 59	0.012
7	60 - 69	0.005
8	70 - 79	0.000
9	80 - 89	0.000
10	90 - 99	0.002
11	Total n	1.000

will open that is divided into five sections (all of the analyses in the LFD3 Add-in will look similar). The program that completes the frequency distribution is a complicated program—to take you through every option would take many pages and figures. Some of the options may be self-explanatory; others may be confusing. Don't worry—we've provided very detailed help with the add-in. In other words, if you don't know what "Grouped By" means, click on the Help button and find out!

By entering the appropriate range of data, starting value, and interval width, you should be able to produce something like Figure 2.8.

As you can see, running the frequency distribution add-in quickly summarizes a large data set in both tabular and graphical form. The graph of the YRSMK data is especially revealing: The distribution has a positive skew with a central tendency in the interval 20–29.

However, like any convenient procedure, the results produced by an add-in can be inappropriate and deceiving. Interpreting the results of an analysis that has not been well thought out (What should the class interval size be? What should the starting value be?) may be difficult or ineffective. For example, if you look closely at the frequency distribution in Figure 2.8, the interval 90–99 contains a proportion of .002 of the scores. The proportion .002 represents a single score out of the 608 total scores (rounded up). Therefore, there is a single person in the Smoking Study who has smoked for 90 or more years (92 years, actually). Is it possible that a person in the study has been smoking for 92 years? If so, they must be at least 100 years old! Upon closer inspection of the data set, though, you can see that this participant is 45 years old. In this case, it appears that there was a data entry error (it should read 29, not 92).

Although this is a relatively minor mistake, which was subsequently caught by the experimenters, its impact on our frequency distribution shouldn't be overlooked. Note that the two preceding intervals, 70–79 and 80–89, are empty (contain a 0.000 proportion of the scores). Thus, our "summary" of this data set may have too few intervals, thereby obscuring some potentially interesting features (if we removed the last three intervals, there would be seven remaining, which, according to the guidelines is too few).

Estimating Percentile Ranks with Excel

In Table 2.2, each of 20 children was given an "aggressiveness score" from 0 to 5. Let's say that you are interested in the percentile rank of a score of 3. Using Excel to estimate the percentile rank of the aggressive score "3," use the worksheet function "PERCENTRANK," which *requires* two arguments, "ARRAY" and "X." If you enter appropriate values for these arguments, Excel will estimate the percentile rank of the value X in the distribution of values specified by ARRAY. For this function to work, ARRAY has to be a "Range" in Excel, or a group of cells that contain numbers. X also has to be a number, although it doesn't have to be a number in the ARRAY.

By entering all the aggressiveness scores in a single column in an Excel spreadsheet and using the PERCENTRANK function, the value "0.473" appears. Multiplying by 100% = 47.30%. In other words, the percentile rank of the score of 3 in the aggressiveness scores is 47.3%.

Estimating Percentiles

Lastly, what is the score value with *P%* of the measurements below it, or the *P*th percentile? Using Excel's "PERCENTILE" function, the 35th percentile of YRSMK, for example, requires entering the range of data and percentile as arguments in the function. Doing this, Excel returns a value 19. In other words, 35% of the scores are at or smaller than 19. Put another way, a score of 19 is the 35th percentile.

Summary

The purpose of descriptive statistics is to organize and summarize data without distortion. One way of achieving this goal is to construct a frequency distribution or a relative frequency distribution. A frequency distribution is a tally of the number of times a particular score value appears in a sample or population. Relative frequency is obtained by dividing these tallies by the total number of measurements. Because relative frequency combines information about frequency and the total number of scores, it is generally more useful than raw frequency. Cumulative frequency (and cumulative relative frequency) distributions tally the number of occurrences of measurements at or below a given score value. They are most useful in calculating percentiles and percentile ranks.

When the measurements in a distribution are scattered across a large number of different score values, frequencies are tabulated for class intervals rather than for individual score values. Choice of interval size is not automatic; the intervals must be chosen so that they do not distort the data. Although there are heuristic suggestions for choosing interval size, the ultimate decision rests on your analysis of the specific distribution with which you are working.

Graphs of frequency distributions highlight major characteristics of distributions and facilitate comparison among distributions. Histograms use the heights of bars to indicate frequency or relative frequency and are often used for nominal scales and when the

measured variable is discrete. Polygons use connected points to indicate frequency and relative frequency. Polygons are often used for measurements of continuous variables. Most commonly, distributions are described and compared in terms of shape (whether the distribution is symmetric or skewed and modality), central tendency (typical score value), and variability (degree to which the scores differ from each other).

Percentile ranks are a measure of relative standing. The percentile rank of a score value indicates the percent of measurements in the distribution below that score value. A score value with a percentile rank of *P*% is called the *P*th percentile. Reporting the percentile rank of a score provides much more information than the score value alone, because the percentile rank indicates the position of that score within the distribution.

Nonetheless, three precautions should be used when interpreting percentile ranks. First, percentile ranks are only approximate. Second, percentile ranks are an ordinal index of relative standing, even when the original measurements are interval or ratio. Third, interpretation of percentile ranks depends on the particular sample (or population) that is described by the distribution.

exercises

terms *Define these new terms.*

frequency distribution
relative frequency
cumulative frequency
cumulative relative frequency
class interval
midpoint
lower real limit
symmetric distribution
positive skew
negative skew
unimodal
bimodal
upper real limit
abscissa
ordinate
histogram
frequency polygon
discrete variable
continuous variable
central tendency
variability
percentile rank
percentile

Questions *Answer the following questions.*

1. Table 2.9 contains three sets of scores. For each set construct
   a. frequency and relative frequency distributions.
   b. a cumulative relative frequency distribution.
2. Assume that the first set of scores in Table 2.9 represents the IQ scores of children who have been participating in a school lunch program, and the second set of scores represent the IQs of a similar group of children who have not participated in the program.