BOOKS | À LA CARTE EDITION

SIXTH EDITION

Statistics for the Behavioral and Social Sciences

A Brief Course

Arthur Aron
Elliot J. Coups
Elaine N. Aron

YOUR TEXTBOOK—IN A BINDER-READY EDITION!

This unbound, three-hole-punched version of your textbook lets you take only what you need to class and incorporate your own notes—all at an affordable price!

SIXTH EDITION

Statistics for the Behavioral and Social Sciences

A Brief Course

Arthur Aron
State University of New York at Stony Brook

Elliot J. Coups
Rutgers Robert Wood Johnson Medical School
Rutgers, The State University of New Jersey

Elaine N. Aron
State University of New York at Stony Brook

Pearson

330 Hudson Street, NY NY 10013

Portfolio Manager: *Tanimaa Mehra*
Content Producer: *Kani Kapoor*
Portfolio Manager Assistant: *Anna Austin*
Product Marketer: *Jessica Quazza*
Art/Designer: *Integra Software Services Pvt. Ltd.*
Digital Studio Course Producer: *Ruth Berry*
Full-Service Project Manager: *Integra Software Services Pvt. Ltd.*
Compositor: *Integra Software Services Pvt. Ltd.*
Printer/Binder: *LSC Communications, Inc./Owensville*
Cover printer: *Phoenix Color*
Cover design: *Lumina Datamatics, Inc.*
Cover Art: *Shutterstock*

Acknowledgments of third party content appear on pages within the text.

Library of Congress Cataloging-in-Publication Data

Names: Aron, Arthur, author. | Coups, Elliot J., author. | Aron, Elaine, author.
Title: Statistics for the behavioral and social sciences: a brief course/Arthur Aron, Ph.D., State University of New York at Stony Brook, Elliot J. Coups, Ph.D., Rutgers Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Elaine N. Aron, Ph.D., State University of New York at Stony Brook.
Description: Sixth Edition. | Hoboken: Pearson, [2017] | Series: Books a la carte | Revised edition of the authors' Statistics for the behavioral and social sciences, c2011.
Identifiers: LCCN 2017050636| ISBN 9780205989065 | ISBN 0205989063
Subjects: LCSH: Social sciences—Statistical methods. | Social sciences—Data processing.
Classification: LCC HA29 .A745 2017 | DDC 300.1/5195—dc23
LC record available at https://lccn.loc.gov/2017050636

1 18

Pearson

Books a la Carte
ISBN–10: 0-20-598906-3
ISBN–13: 978-0-20-598906-5

Access Code Card
ISBN–10: 0-13-487719-5
ISBN–13: 978-0-13-487719-8

Instructor's Review Copy
ISBN–10: 0-13-441933-2
ISBN–13: 978-0-13-441933-6

Rental
ISBN–10: 0-13-489902-4
ISBN–13: 978-0-13-489902-2

Brief Contents

Contents

Preface to the Instructor

The heart of this text was written over a summer in a small apartment near the Place Saint Ferdinand, having been outlined in nearby cafes and on walks in the Bois de Boulogne. It is based on our many years of experience in teaching, researching, and writing. We believe that it is still as different from the conventional lot of statistics texts as Paris is from Pompeii, yet still comfortable and stimulating to the long-suffering community of statistics instructors.

Our approach was developed over decades of successful teaching—successful not only in the sense that students have consistently rated the course (a statistics course, remember) as a highlight of their undergraduate years, but also in the sense that students come back to us later saying, "I was light-years ahead of my fellow graduate students because of your course," or "Even though I don't do research, your course has really helped me understand statistics that I read about in my field."

In this sixth edition of this *Brief Course* we have tried to maintain those things about the course that have been especially appreciated, while reworking the text to take into account the feedback we have received, our own ongoing teaching experiences, and advances and changes in the field. We of course have continued to focus on simplifying exposition and have done our usual updating of content, examples, and so on, plus making a host of minor adjustments to make the book more effective. For several chapters, our examples are now even more engaging, which we hope will further increase student interest and learning.

New to this Edition

As we undertook the challenge to update and refresh the sixth edition of our text, some of the major changes include:

- We have added learning objectivs that tie to each section and summary point in each chapter. This allows students to identify and track their understanding of key concepts.
- New, recent examples that illustrate each statistical procedure in actual research articles.
- More subsections, or "chunking" of information to help your students better absorb and retain statistical concepts and their applications throughout this course.
- Appendix E, A Closer Look at the Logic and Language of Behavioral and Social Sciences Research, formerly Web Chapter 1. We have updated this chapter (and include it now in with the main text, making it more easily available), which helps students to understand and evaluate the research contexts in which statistical methods are applied.
- Chapter 12, A Closer Look at Advanced Statistical Procedures in Research Articles, formerly Web Chapter 2 (and also previously not included in the main text), which helps students to make sense of advanced statistical procedures that they will commonly encounter in research articles. This chapter has been updated to reflect current techniques, such as more recent procedures for conducting mediation analyses and the basic ideas of Bayesian methods.

In addition to the new components of our text, we also want to reiterate some comments we made in the first edition about how this book from the beginning has been quite different from other statistics texts.

What We Do Differently

As different as this course is, it has from the start also done what the best of the statistics texts of the last few years have been already doing well: emphasizing the intuitive, deemphasizing the mathematical, and explaining everything in direct, simple language. But what we have done continues to go beyond even the best of the current lot in eight key respects.

1. ***The definitional formulas are brought to center stage*** because they provide a concise symbolic summary of the logic of each particular procedure. All our explanations, examples, practice problems, and test bank items are based on these definitional formulas. (The amount of data to be processed in our practice problems and test items are reduced appropriately to keep computations manageable.)

   Why this approach? Even in 2018 most statistics texts have still not faced technological realities. What is important is not that the students learn to calculate a correlation coefficient by hand with a large data set—programs like SPSS can do this in an instant with just a few mouse clicks. What is important is that students work problems in a way that keeps them constantly aware of the underlying logic of what they are doing. Consider the population variance—the average of the squared deviations from the mean. This concept is immediately clear from the definitional formula (once the student is used to the symbols): the population variance $= [\Sigma(X - M)^2]/N$. Repeatedly working problems using this formula ingrains the *meaning* in the student's mind. In contrast, the usual computational version of this formula only obscures this meaning: the population variance $= [\Sigma X^2 - (\Sigma X)^2/N]/N$. Repeatedly working problems using this computational formula does nothing but teaches the student the difference between ΣX^2 and $(\Sigma X)^2$!

Teaching these tired computational formulas today is an anachronism—at least 40 years out of date! Researchers do their statistics on computers, and the use of statistical software makes the understanding of the basic principles, as they are symbolically expressed in the definitional formula, more important than ever. Students still need to work lots of problems by hand to learn the material. But they need to work them using the definitional formulas that reinforce the concepts, not using the antiquated computational formulas that obscure them. Not since the era when Eisenhower was U.S. president have those computational formulas made some sense as time-savers for researchers who had to work with large data sets by hand. But they were always poor teaching tools. (Because some instructors may feel naked without them, we still provide the computational formulas, usually in a brief footnote, at the point in the chapter where they would traditionally have been introduced.)

2. ***Each procedure is taught both verbally and numerically—and usually visually as well.*** In fact, when we introduce *every* formula, it has attached to it a concise statement of the formula in words. (The major formulas *with their verbal descriptions* are also repeated in Appendix D.) Typically, each example lays out the procedures in worked-out formulas, in words (often with a list of steps), and usually illustrated with easy-to-grasp figures. Practice problems and test bank items, in turn, require the student to calculate results, write a short explanation in layperson's language of what they have done, and make a sketch (for example, of the distributions involved in a *t* test). The course material completely prepares the student for these kinds of practice problems and test questions.

   It is our repeated experience that these different ways of expressing an idea are crucial for permanently establishing a concept in a student's mind. Many students in the behavioral and social sciences are more at ease with words than with numbers. In fact, many of even the brightest students have a strong fear of all mathematics. Writing the formula in words and providing the lay-language explanation give them an opportunity to do what they do best.

3. A main goal of any introductory statistics course in the behavioral and social sciences is to ***prepare students to read research articles.*** The way a procedure such as a *t* test or chi-square is described in a research article is often quite different from what the student expects from the standard textbook discussions. Therefore, as this course teaches a statistical method, it also gives examples of how that method is reported in recent journal articles. And we don't just leave it there. The practice problems and test bank items also include excerpts from journal articles for the student to explain.

4. The book is ***unusually up-to-date.*** For some reason, most introductory statistics textbooks read as if they were written in the 1950s. The basics are still the basics, but statisticians and researchers think far more subtly about those basics now. Today, the basics are undergirded by a new appreciation of issues such as effect size, power, the accumulation of results through meta-analysis, the critical role of models, and a whole host of new orientations arising from the central role of the computer in statistical analyses. We are much engaged in the latest developments in statistical theory and application, and this book reflects that engagement. For example, we devote an entire early chapter to effect size and power and then return to these topics as we teach each technique. Furthermore, we discuss how to handle situations in which assumptions are violated, and we cover data transformations (this widely used approach is easily accessible to introductory students but is rarely mentioned in current introductory texts).

5. ***We capitalize on the students' motivations.*** We do this in two ways. First, our examples, while attempting to represent the diversity of behavioral and social sciences research, emphasize topics or populations that students seem to find most interesting. The very first example is from a real study in which students in their first week of an introductory statistics class rate how much stress they feel they are under. Also, our examples continually emphasize the usefulness of statistical methods and ideas as tools in the research process, never allowing students to feel that what they are learning is theory for the sake of theory.

   Second, we have worked to make the course extremely straightforward and systematic in its explanation of basic concepts so that students can have frequent "aha!" experiences. Such experiences bolster self-confidence and motivate further learning. It is quite inspiring to *us* to see even fairly modest students glow from having mastered some concept like negative correlation or the distinction between failing to reject the null hypothesis and supporting the null hypothesis. At the same time, we do not constantly remind them how greatly oversimplified we have made things, as some textbooks do.

6. ***We emphasize statistical methods as a living, growing field of research.*** Each chapter includes a "Bringing Statistics to Life" feature about famous statisticians or interesting sidelights about the field. The goal is for

students to see statistical methods as human efforts to make sense out of the jumble of numbers generated by a research study; to see that statistics are not "given" by nature, not infallible, not perfect descriptions of the events they try to describe, but rather constitute a language that is constantly improving through the careful thought of those who use it. We hope that this orientation will help them maintain a questioning, alert attitude as students and later as professionals.

7. ***We include a chapter-length appendix (Appendix E) that provides an overview of the logic and language of the research process in behavioral and social sciences.*** Statistical methods are tools used in the research process. Students will find the statistical procedures in this course easier to understand if they appreciate the context in which they are typically used. This chapter outlines the pros and cons of numerous common research designs and also introduces students to the important topics of reliability and validity.
8. ***We include a chapter (Chapter 12) that looks at advanced procedures*** without actually teaching them in detail. It explains in simple terms how to make sense out of these statistics when they are encountered in research articles. Most research articles today use methods such as multilevel modeling, mediation, factor analysis, structural equation modeling, analysis of covariance, or multivariate analysis of variance. Students completing an ordinary introductory statistics course are ill-equipped to comprehend most of the articles they must read to prepare a paper or study a course topic in further depth. This chapter makes use of the basics that students have just learned (along with extensive excerpts from recent research articles) to give a rudimentary understanding of these advanced procedures. This chapter also serves as a reference guide that students can use in the future when reading such articles.
9. ***We include a chapter (Chapter 13) that helps sudents learn how to apply statistical methods in their own research projects.*** The topics covered include: determining the type of statistical test to use; determining statistical power and required sample size; gathering the data for a study; screening data before analysis; implementing a systematic analysis plan; and writing up the study results. The chapter helps students to understand the broader application of the statistical principles and tests that they learned in prior chapters.

Support Materials for Instructors

This book is also different from most others in the amount of support materials for instructors (and in this edition we have further strengthened the instructor's resource materials to make the course easier than ever to teach).

INSTRUCTOR'S MANUAL (ISBN 9780205989126)

We have written an ***Instructor's Manual that really helps teach the course.*** The manual begins with a chapter summarizing what we have gleaned from our own teaching experience and the research literature on effectiveness in college teaching. The next chapter discusses alternative organizations of the course, including tables of possible schedules and a sample syllabus. A lecture outline is provided for each chapter, along with teaching examples (beyond those in the text) that can be easily incorporated into your lectures. The manual also includes computational answers to each chapter's Set 2 practice problems. (The text provides answers to the Set 1 practice problems at the back of the book, including at least one example answer to an essay-type question for each chapter.)

TEST BANK (ISBN 9780205989126)

Our ***Test Bank section of the Instructor's Manual makes preparing good exams easy.*** We supply approximately 40 multiple-choice, 25 fill-in, and 10–12 problem/essay questions for each chapter. Considering that the emphasis of the course is so conceptual, the multiple-choice questions will be particularly useful for those of you who do not have the resources to grade essays.

POWERPOINT SLIDES (ISBN 9780205989140)

A set of customizable ADA-compliant PowerPoint slides is now available for each chapter. Together with the Instructor's Manual, you now have a wealth of resources that with minimal time will help you to prepare and deliver a stellar lecture, every single class.

About this *Brief Course*

We have been thrilled by the enthusiastic response of instructors and students to the six editions of our *Statistics for Psychology* (Aron & Aron, 1994, 1999, 2003; Aron, Aron, & Coups, 2006, 2009; Aron, Coups, & Aron, 2013), as well as the positive comments of reviewers, including most encouraging evaluations from early on in *Contemporary Psychology* (Bourgeois, 1997) and *Psychology Learning and Teaching* (Shevlin, 2005).

This *Brief Course* was our answer to the many requests we received from instructors and students for a textbook using our approach that is (a) more general in its focus than psychology alone and (b) shorter, to accommodate less comprehensive courses. Of course, we tried to retain all the qualities that endeared the original to our readers. At the same time, the *Brief Course* was not a cosmetic revision. The broadening of focus meant using examples from the entire range of behavioral and social sciences, from anthropology to political science. Most important, the broadening informed the relative emphasis (and inclusion) of different topics and the tenor of the discussion of these topics. The shortening was also dramatic:

This *Brief Course* is substantially briefer than the original, making it quite feasible to cover the whole text even in a quarter-length course.

Keep in Touch

Our goal is to do whatever we can to help you make your course a success. If you have any questions or suggestions, please send us an e-mail (Arthur.Aron@stonybrook.edu will do for all of us). Also, if you should find an error somewhere, for everyone's benefit, please let us know right away. When errors have come up in the past, we have usually been able to fix them in the very next printing.

Acknowledgments

First and foremost, we are grateful to our students through the years, who have shaped our approach to teaching by rewarding us with their appreciation for what we have done well, as well as their various means of extinguishing what we have done not so well. We also much appreciate all those instructors who have sent us their ideas and encouragement.

We remain grateful to all of those who helped us with the first five editions of this book, as well as to those who helped with the first six editions of the larger book (*Statistics for Psychology*). We thank the many people at Pearson who guided and supported us through the revision process for this sixth edition of the *Brief Course*.

Arthur Aron
Elliot J. Coups
Elaine N. Aron

Introduction to the Student

The goal of this book is to help you *understand* statistics. We emphasize meaning and concepts, not just symbols and numbers.

This emphasis plays to your strength. Many behavioral and social science students are not lovers of mathematics but are keenly attuned to ideas. And we want to underscore the following, based on our collective many decades of experience in teaching: *We have never had a student who could do well in other university courses who could not also do well in this course*. (However, we will admit that doing well in this course may require more work than doing well in others.)

In this introduction, we discuss why you are taking this course and how you can gain the most from it.

Why Learn Statistics, Other Than to Fulfill a Requirement?

1. ***Understanding statistics is crucial to being able to read research articles.*** In most of the behavioral and social sciences, nearly every course you take will emphasize the results of research studies, and these usually include statistics. If you do not understand the basic logic of statistics—if you cannot make sense of the jargon, the tables, and the graphs that are at the heart of any research report—your reading of research will be very superficial. We also recommend that you take a course on how to design and evaluate good research. We have, however, added an appendix (Appendix E) which provides an overview of the logic and language of the research process in behavioral and social sciences. Also, we have included a chapter (Chatper 12) which helps you understand the basics of the most common more advanced procedures you may run into when reading research, without actually teaching them in detail.
2. ***Understanding statistics is crucial to doing your own research.*** Many students eventually go on to graduate school. Graduate study in the behavioral and social sciences almost always involves *doing* research. In fact, learning to do research on your own is often the entire focus of graduate school, and doing research almost always involves statistics. This course gives you a solid foundation in the statistics you need for doing research. Further, by mastering the basic logic and ways of thinking about statistics, you will be unusually well prepared for advanced courses, which focus on the nitty-gritty of analyzing research results.

   Many universities also offer opportunities for undergraduates to do research. The main focus of this course is understanding statistics, not using statistics. Still, you will learn the basics you need to analyze the results of the kinds of research you are likely to do. And in Chapter 13, Applying Statistical Methods in Your Own Research Project, you will learn some practical advice and tips for using what you learn in this text for analyzing the results of your own research.
3. ***Understanding statistics develops your analytic and critical thinking.*** Behavioral and social science students are often most interested in people and in improving things in the practical world. This does not mean that you avoid abstractions. In fact, the students we know are exhilarated most by the almost philosophical levels of abstraction where the secrets of human experience so often seem to hide. Yet even this kind of abstraction often is grasped only superficially at first, as slogans instead of useful knowledge. Of all the courses you are likely to take in the behavioral and social sciences, this course will probably do the most to help you learn to think precisely, to evaluate information, and to apply logical analysis at a very high level. Moreover, you will find that these skills will also help you to evaluate media reports of behavioral and social sciences research, which are becoming increasingly common.

How to Gain the Most from This Course

There are five things we can advise:

1. ***Keep your attention on the concepts.*** Treat this course less like a math course and more like a course in logic. When you read a section of a chapter, your attention should be on grasping the principles. When working the exercises, think about why you are doing each step. If you simply try to memorize how to come up with the right numbers, you will have learned very little of use in your future studies—nor will you do very well on the tests in this course.
2. ***Be sure you know each concept before you go on to the next.*** Statistics is cumulative. Each new concept is built on the last one. There are short "How are you doing?"

self-tests at the end of each main chapter section. Be sure you do them. And if you are having trouble answering a question—or even if you can answer it but aren't sure you really understand it—stop. Reread the section, rethink it, ask for help. Do whatever you need to do to grasp it. Don't go on to the next section until you are completely confident you have gotten this one. If you are not sure, and you've already done the "How are you doing?" questions, take a look at the "Example Worked-Out Problems" toward the end of the chapter, or try working a practice problem on this material from the end of the chapter. The answers to the Set 1 practice problems are given in Appendix B at the end of the book, so you will be able to check your work.

Having to read the material in this text over and over does not mean that you are stupid. Most students have to read each chapter several times. And each reading in statistics is usually much slower than that in other textbooks. Statistics reading has to be pored over with clear, calm attention for it to sink in. Allow plenty of time for this kind of reading and rereading.

3. ***Keep up.*** Again, statistics is cumulative. If you fall behind in your reading or miss lectures, the lectures you then attend will be almost meaningless. It will get harder and harder to catch up.
4. ***Study especially intensely in the first half of the course.*** It is especially important to master the material thoroughly at the start of the course. Everything else you learn in statistics is built on what you learn at the start. Yet the beginning of the semester is often when students study least.

   If you have mastered the first half of the course—not just learned the general idea, but really know it—the second half will be a lot easier. If you have not mastered the first half, the second half will be close to impossible.
5. ***Help each other.*** There is no better way to solidify and deepen your understanding of statistics than to try to explain it to someone having a harder time. (Of course, this explaining has to be done with patience and respect.) For those of you who are having a harder time, there is no better way to work through the difficult parts than by learning from another student who has just mastered the material.

   Thus, we strongly urge you to form study groups with one to three other students. It is best if your group includes some who expect this material to come easily and some who don't. Those who learn statistics easily will get the very most from helping others who struggle with it—the latter will tax the former's supposed understanding enormously. Those who fear trouble ahead, you need to work with those who do not—the blind leading the blind is no way to learn. Pick group members who live near you so that it is easy for you to get together. Also, meet often—between each class, if possible.

A Final Note

Believe it or not, we love teaching statistics. Time and again, we have had the wonderful experience of having beaming students come to us to say, "Professor, I got a 90% on this exam. I can't believe it! Me, a 90 on a statistics exam!" Or the student who tells us, "This is actually fun. Don't tell anyone, but I'm actually enjoying statistics, of all things!" We very much hope you will have these kinds of experiences in this course.

Arthur Aron
Elliot J. Coups
Elaine N. Aron

Chapter 1

Displaying the Order in a Group of Numbers Using Tables and Graphs

Learning Objectives

1.1 Outline the use of statistical concepts in behavioral and social sciences

1.2 Differentiate between descriptive and inferential statistics

1.3 Define basic statistical concepts

1.4 Analyze data using frequency tables

1.5 Create histograms from frequency tables

1.6 Identify the shape of a distribution of data

1.7 Interpret shapes of distributions from their verbal descriptions in research articles

Welcome to *Statistics for the Behavioral and Social Sciences: A Brief Course*. We imagine you to be as unique as the other students we have known who have taken this course. Some of you are highly scientific sorts; others are more intuitive. Some of you are fond of math; others are less so. Whatever your style, we welcome you. We want to assure you that if you give this course some special attention, you *will* learn statistics. The approach we use has successfully taught all sorts of students before you, including those who had taken statistics previously and done poorly. As you read through these chapters, and with your instructor's help, you will learn statistics and learn it well.

Given that you *can* learn statistics, you still have to decide if you want to make the effort it will require. Why would you want to do that, except to meet a requirement of your major? (Not a very energizing motive.)

1. You will be far better equipped to read research articles in your major.
2. You will be on your way to being able to do your own research if you so choose.
3. You will improve both your reasoning and your intuition, which will help you to understand and evaluate statistics that are reported in the media and books.
4. Each time you finally grasp something you were struggling with, you will feel great.

What Is Statistics?

1.1 Outline the use of statistical concepts in behavioral and social sciences

Formally, *statistics* is a branch of mathematics that focuses on the organization, analysis, and interpretation of a group of numbers. But really what is statistics? Think of statistics as a tool that extends a thinking process used by every human: you observe a thing; you wonder what it means or what caused it; you have an insight or make an intuitive guess; you observe again, but now in detail, or you try making some little changes in the process to test your intuition. Then you face the eternal problem: Was your hunch confirmed or not? What are the chances that what you observed this second time will happen again and again so that you can announce your insight to the world as something probably true?

Statistics is a method of pursuing truth. At a minimum, statistics can tell you the likelihood that your hunch is true at this time and place and with these sorts of people. This pursuit of truth, or at least its future likelihood, is the essence of science. It is also the essence of human evolution and survival. Think of the first research questions: What will the mammoths do next spring? What will happen if I eat this root? It is easy to see how the early accurate "researchers" survived. You are here today because your ancestors exercised brains as well as brawn. Do those who come after you the same favor: Think carefully about outcomes. Statistics is one good way to do that.

Behavioral and social scientists use statistical methods to help them make sense of the numbers they collect when conducting research. They will usually use a computer and statistical software to carry out statistical procedures such as the ones you will learn here. However, the best way to develop a solid understanding of statistics is to learn how to do the procedures by hand (with the help of a calculator—it's not the multiplying and adding that you learn from, but the going through all the steps). In order to minimize the amount of figuring you have to do, we use relatively small groups of numbers in each chapter's examples and practice problems. We hope that this will also allow you to focus more on the *underlying principles and logic* of the statistical procedures, rather than on the computations for each practice problem (such as subtracting 3 from 7 and then dividing the result by 2 to give an answer of 2).

Having said that, our emphasis is on you being able to see the logic of the procedures by working with numerically simple examples. We also recognize the importance of learning how to do statistical procedures on a computer that can handle complex data, as you most likely would when conducting your own research. Thus, at the end of relevant chapters, there is a section called "Using SPSS." SPSS statistical software is commonly used by behavioral and social scientists to carry out statistical analyses. Check with your instructor to see if you have access to SPSS at your institution. (There are also a number of other statistical programs used by researchers; for those of you using such a program with this course, the SPSS sections will still be helpful in giving you the general idea of how to work out problems using computer software.)

The Two Branches of Statistical Methods

1.2 Differentiate between descriptive and inferential statistics

There are two main branches of statistical methods:

1. **Descriptive statistics**: Behavioral and social scientists use descriptive statistics to summarize and *describe* a group of numbers from a research study.
2. **Inferential statistics**: Behavioral and social scientists use inferential statistics to draw conclusions and make *inferences* that are based on the numbers from a research study but that go beyond these numbers. For example, inferential statistics allow researchers to make inferences about a large group of individuals based on a research study in which a much smaller number of individuals took part.

Some Basic Concepts: Variables, Values, and Scores

1.3 Define basic statistical concepts

As part of a study (Aron, Paris, & Aron, 1995), researchers gave a questionnaire to 151 students in an introductory statistics class during the first week of class. One question asked was, "How stressed have you been in the last 2½ weeks, on a scale of 0 to 10, with 0 being not at all stressed and 10 being as stressed as possible?" (How would you answer?) In this study, the researchers used a survey to examine students' level of stress. Some other methods that researchers use to study stress include creating stress with laboratory tasks (like having to be videotaped giving a talk for humans, or swimming in water for rats) and measuring stress-related hormones or brain changes.

In this example, level of stress is a *variable*, which can have *values* from 0 to 10, and the value of any particular person's answer is the person's **score**. Thus, if you answered 6, your score is 6; your score has a value of 6 on the variable called level of stress.

More formally, a variable is a condition or characteristic that can have different values. In short, it can *vary*. In our example, the variable is level of stress. It can have values of 0 through 10. Height is a variable, social class is a variable, score on a creativity test is a variable, number of people absent from work on a given day is a variable, dosage of a medication is a variable, political party preference is a variable, and class size is a variable.

A value is just a number, such as 4, −81, or 367.12. A value can also be a category, such as female or male or the country in which you live (Canada, the United States, Australia, Spain, etc.).

Finally, on any variable, each person has a particular number or score that is that person's value on the variable. For example, your score on the stress variable might have a value of 6. Another student's score might have a value of 8.

Behavioral and social science research is about variables, values, and scores (see Table 1–1). We use these terms throughout the course. The formal definitions are a bit abstract. In practice, you will find that what we mean when we use these words is usually clear.

To learn about a smaller, informal study that these researchers did regarding students' statistics anxiety, and how students who have this anxiety lessen it, read Bringing Statistics to Life Box 1–1.

1.3.1 Kinds of Variables

Most of the variables that behavioral and social scientists use are like those in the stress ratings example: The scores are numbers that tell you how much there is of the thing being measured. In the stress ratings example, the higher the number is, the more stress there is. We call this kind of variable a *numeric variable*. Numeric variables are also called *quantitative variables*.

Behavioral and social scientists use two main kinds of numeric variables: equal-interval variables and rank-order variables. The kind of variable used most often is a variable in which the numbers stand for approximately equal amounts of what is being measured. This is called an *equal-interval variable*. Take grade point average (GPA): This is a roughly equal-interval variable. For example, the difference between a GPA of 2.5 and 2.8 means about as much of a difference as the difference between that of 3.0 and 3.3 (each is a difference of 0.3 of a GPA). Most behavioral and social scientists also consider scales like the 0–10 stress ratings as roughly equal interval. So, for example, a difference between stress rating of 4 and 6 means about as much as the difference between 7 and 9.

Table 1–1 Some Basic Terminology

Term	Definition	Examples
Variable	Condition or characteristic that can have different values	Stress level; age; gender; religion
Value	Number or category	0, 1, 2, 3, 4; 25, 85; female; Catholic
Score	A particular person's value on a variable	0, 1, 2, 3, 4; 25, 85; female; Catholic

BRINGING STATISTICS TO LIFE

Box 1–1 Statistics Anxiety, Test Anxiety, and You

Feeling nervous about this course? Most of you probably are and your instructor knows it too. When one of us was speaking on teaching statistics to over 100 instructors at a conference, we used a questionnaire to ask, "What's your greatest problem teaching stats?" By far the most common answer had to do with their students' anxiety about the course. We need to tackle this right now.

Math Anxiety from Your Past or Statistics Anxiety in Your Present?

Math anxiety is the subject of many good books and studies, and now "statistics anxiety" is as well. We can reassure you on two points: First, this course is a chance for a fresh start with digits, and your past performance in (or avoidance of) geometry, trigonometry, calculus, or similar horrors need not influence in any way how well you comprehend statistics, a largely different subject. If you did okay in basic high school algebra, you will be fine. Forgotten it? You can review it.

The standard test of statistics anxiety (Cruise & Wilkins, 1980) identifies several problems—mainly, the level of interest in statistics and the perceived value of learning it, in addition to the fear of the computations and of understanding results. Although only anecdotal evidence, many former students have approached us over the years saying how this course was different because of this book. They said the book made statistics far less difficult, and more interesting, understandable, and valuable than they expected. Some even conquered their fear enough to pursue graduate work, instead of ruling out higher level courses because of the required statistics.

Even with our reassurances, math anxiety may still be an issue for you. A web search for "mathematics anxiety" will yield many excellent resources. But here's what we have seen help statistics-anxious students when we teach it ourselves:

- Go to class faithfully and stay alert throughout its entirety.
- Do your statistics homework before you are tired.
- Ask for help from your instructor, teaching assistants, or anyone you know in the class.
- Form a study group. If you don't understand something, try explaining it to someone who understands it even less.
- Read ahead, so that what you hear in class is already familiar.
- Do not expect to understand things instantly; treat yourself gently.

Test Anxiety

Now let's focus on exams. You may feel you will understand the material, but still fear not being able to demonstrate this on tests. Anxiety produces physiological arousal; in fact, one of the best understood relationships in psychology is between arousal and performance. Whereas moderate arousal helps performance, including performance on statistics exams (Keeley, Zayac, & Correia 2008) and your overall grade (Macher, Paechter, Papousek, & Ruggeri, 2012; Macher, Paechter, Papousek, Ruggeri, Freudenthaler, & Arendasy, 2013), too much (or too little) test anxiety dramatically decreases cognitive capacity and working memory.

What can you do? One of the leading researchers in the field of test anxiety, Sian Beilock (2010), has written a book on this, *Choke,* which validates our own experiences. Recommended strategies include the following:

1. Overprepare until it seems you cannot possibly fail, even when anxiety has reduced your working memory. Overprepare especially for the first test, when there is no old material to review. A good score on it will make you less anxious about the next test.
2. Practice working under pressure. Do a practice test of homework problems, making it as similar to a real test as possible. Duplicate the most bothersome aspects by setting a time limit. Having to write answers fully and legibly can make you feel slow during a test, so do that too. If the presence of other people bothers you, do a practice test with others in your course. You can even make it a competition, to practice accepting that you may not finish first or score highest.
3. Try expressive writing about your anxiety just before the test. Quite paradoxically, this helps math-anxious students perform as well as low-anxious students, and better than a control group sitting quietly just before the test (Park, Ramierez, & Beilock, 2014). Specifically, about ten minutes before the test, write about how you feel. Be very explicit about your anxiety, what is it is all about, and other times you have felt it. Be honest. No one will see it.
4. Look for which sections are given the most points during the test. If these are not difficult for you, begin with these. Otherwise start with the questions easiest for you. Read instructions carefully! On a multiple-choice exam, check the total number of questions so you can pace yourself. Skip questions you are not sure about, maybe noting the most likely answer so, if necessary, you can guess at the end and have an answer for everything. Try to ignore the people around you. Some people may seem to be working faster, but they may not be working as well as you.

Is your problem general anxiety or general lack of confidence? Then we suggest that you visit your college counseling center.

Lastly, could you be highly sensitive? High sensitivity is an innate trait found in about 20% of humans and is mostly a very useful characteristic, but a downside is being easily overaroused under pressure. You can find out if you are highly sensitive at http://www.hsperson.com. If you are, appreciate the trait's assets, which will definitely help you throughout your life, and make some allowances for its one disadvantage (overarousal).

The other kind of numeric variable behavioral and social scientists often use refers to where the numbers stand in their relative rankings. This is called a *rank-order variable*. (Rank-order variables are also called *ordinal variables*.) Where students stand in their graduating class is one such example. Notice that with a rank-order variable, the difference between one number and the next does not always mean the same amount of the underlying thing being measured. For example, the difference between being second and third in your graduating class could be a very unlike amount of difference in underlying GPA than the difference between being eighth and ninth. A rank-order variable provides less information than an equal-interval variable. It is less precise. Behavioral and social scientists usually use rank-order variables when this is the only information available.

There is also a kind of variable that is not about numbers at all, but that refers just to categories or names. This is called a *categorical variable* because its values are categories. (Categorical variables are also called *nominal variables* because their values are names.) For example, for the categorical variable "undergraduate major," the values might be psychology, sociology, anthropology, economics, engineering, history, and so on.

Categorical variables can indeed be very important—even for you, learning statistics. For example, gender and ethnicity elicit common stereotypes that often suggest that people of some social categories will tend "by nature" do worse at math (which is of course not true). However, if a person of such a group is reminded in any subtle way of the stereotype before doing a math problem or exam, he or she is more likely to do worse than if not reminded of it. Does this apply to you? See Bringing Statistics to Life 1–2.

BRINGING STATISTICS TO LIFE

Box 1–2 Gender, Ethnicity, and Math Performance

Let's begin with an important concept in social psychology: "stereotype threat." Stereotype threat refers to a situation or setting in which a negative stereotype reminds you of culturally induced doubts about your ability (e.g., in math or statistics) due to your gender, race, ethnicity (Steele, 1997), disability (Silverman & Cohen, 2014), or even older age (Krendl, Ambady, & Kensinger, 2015). Reminders can be very subtle. For example, in one research study, before a math test fifth- and sixth-grade students were randomly assigned to fill out a form asking for their race (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010). African American students who filled out this form answered only about 50% correctly compared to African American students who were not asked to fill out the form! In another study, before a math test women who were told that men generally score better did much worse on the test than women who were told that the two genders score equally well (Spencer, Steele, & Quinn, 1999).

What Can You Do for Yourself?

Of course, you may have legitimate reason to doubt your mathematical abilities because as a woman or minority group member you were not exposed to much math, for whatever reason. Happily, as noted in Bringing Statistics to Life 1–1, this course does not rely very much on past math experience. Brush up on algebra I and you are all set.

As for stereotype threat, the best solution is just to eliminate the negative stereotypes. For example, you were not born "bad at math." Research indicates that prejudices, not genetics or brain power, are the cause of differences in test scores. For example, the same difference of 15 IQ points between a dominant and minority group of any type has been found all over the world, even when there is no genetic difference between the groups, and where opportunities for a group have changed (e.g., due to migration) differences have rapidly disappeared (Block, 1995). The more gender equality in a society, the more that women are equal to men in math performance and the greater likelihood that women are found to have "profound mathematical talent" (Hyde & Mertz, 2009). In fact, overall there are no differences between the genders in math performance (Hyde, 2016).

If your feelings of stereotype threat are still there at exam time, try the method used in the study we mentioned with fifth- and sixth-graders. Stereotype threat was greatly reduced by telling students to think of their test as an opportunity to "learn a lot of new things" and help them in school because taking the test "sharpens the mind." While studying or before an exam, you can do the same: Remember that by taking a test you are learning new things that will sharpen your mind and help you in other courses and throughout your life.

Another very successful method is self-affirmation (Sherman, 2013), in which prior to a test you list your most valued characteristics and write about the highest one on your list, including a time when it was particularly important. It puts your performance in statistics in perspective, doesn't it?

The effect of stereotype threat on arousal, working memory, and simply remembering what you are doing appears to be the same as that of text anxiety (Hutchison, Smith, & Ferris, 2013), so anything that helps test anxiety (see Bringing Statistics to Life 1–1) helps here as well. One study found that during an exam, pushing out of your mind any thoughts created by stereotype threat is not ideal because it requires energy that could otherwise be used for the test (Schuster, Martiny, & Schmader, 2015). It's better to simply acknowledge those feelings as normal and due to stereotypes that are nonsense, and then proceed.

Finally, we encourage you to fight the activation of prejudices, inner or outer, that could stand in the way of your right to know statistics. These are the words from a president of the Mathematics Association of America, written several decades ago, but still true today:

The paradox of our times is that as mathematics becomes increasingly powerful, only the powerful seem to benefit from it. The ability to think mathematically—broadly interpreted—is absolutely crucial to advancement in virtually every career. Confidence in dealing with data, skepticism in analyzing arguments, persistence in penetrating complex problems, and literacy in communicating about technical matters are the enabling arts offered by the . . . mathematical sciences. (Steen, 1987, p. xviii)

1.3.2 Levels of Measuring Variables

Researchers sometimes have to decide how they will measure a particular variable. For example, they might use an equal-interval scale, a rank-order scale, or a categorical scale. These different kinds of variables are based on different *levels of measurement* (see Table 1–2).

The level of measurement selected affects the type of statistics that can be used with a variable. In this course, we focus mostly on numeric equal-interval variables. However, rank-order and categorical variables also are fairly common in the behavioral and social sciences, and we give them attention too.

I. How are you doing?

Answers can be found at the end of this chapter.

1. A father rates his daughter as a 2 on a 7-point scale (from 1 to 7) of crankiness. In this example, (a) what is the variable, (b) what is the score, and (c) what is the range of possible values?
2. What is the difference between a numeric and a categorical variable?
3. Name the kind of variable for each of the following variables: (a) a person's nationality (Mexican, French, Japanese, etc.), (b) a person's score on a standardized IQ test, and (c) a person's place on a waiting list (first in line, second in line, etc.).

Frequency Tables

1.4 Analyze data using frequency tables

In this section, you will learn how to make a table that makes the pattern of numbers in a study easier to see.

1.4.1 Example: Returning to the Stress Rating Study

Let's return to the stress rating example. Recall that in this study, students in an introductory statistics class during the first week of the course answered the question "How stressed have you been in the last 2½ weeks, on a scale of 0 to 10, with 0 being *not at all stressed* and 10 being *as stressed as possible*?" The actual study included scores from 151 students. To ease the learning for this example, we are

Table 1–2 Levels of Measurement

Level	Definition	Example
Equal-interval	Numeric variable in which differences between values correspond to differences in the underlying thing being measured	Stress level, age
Rank-order	Numeric variable in which values correspond to the relative position of things measured	Class standing, position finished in a race
Categorical	Variable in which the values are categories	Undergraduate major, religion

going to use a representative subset of scores from 30 of these 151 students (this also saves you time if you want to try it for yourself). The 30 students' scores (their ratings on the scale) are as follows:

8, 7, 4, 10, 8, 6, 8, 9, 9, 7, 3, 7, 6, 5, 0, 9, 10, 7, 7, 3, 6, 7, 5, 2, 1, 6, 7, 10, 8, 8.

Looking through all these scores gives some sense of the overall tendencies. But this is hardly an accurate method. One solution is to make a table showing how many students used each of the 11 values that the ratings can have (0, 1, 2, etc., through 10). That is, the number of students who used each particular rating is the *frequency* of that value. We have done this in Table 1–3.

We also figured the percentage that each value's frequency is of the total number of scores. Tables like this sometimes give only the raw-number frequencies and not the percentages, or only the percentages and not the raw-number frequencies.[1]

Table 1–3 is called a *frequency table* because it shows how frequently (how many times) each rating number was used. A frequency table makes the pattern of numbers easy to see. In this example, you can see that most of the students rated their stress around 7 or 8, with few rating it very low. How would you rate on this frequency table and why? How do you perceive your own abilities? Read Bringing Statistics to Life Boxes 1–1 and 1–2.

1.4.2 How to Make a Frequency Table

There are four steps in making a frequency table.

❶ **Make a list of each possible value down the left edge, starting from the lowest and ending with the highest.** In the stress rating results, the list goes from 0, the lowest possible rating, up to 10, the highest possible rating. Note that even if one of the ratings between 0 and 10 had not been used, you would still include that value in the listing, showing it as having a frequency of 0. For example, if no one gave a stress rating of 2, you still include 2 as one of the values on the frequency table. (Most research articles follow the procedure we recommend here: Going from lowest at the top to highest at the bottom. However, some statistics authorities prefer to go from highest at the top to lowest at the bottom.)

❷ **Go one by one through the scores, making a mark for each next to its value on your list.** This is shown in Figure 1–1.

❸ **Make a table showing how many times each value on your list was used.** To do this, add up the number of marks beside each value.

Table 1–3 Frequency Table from the Stress Rating Example

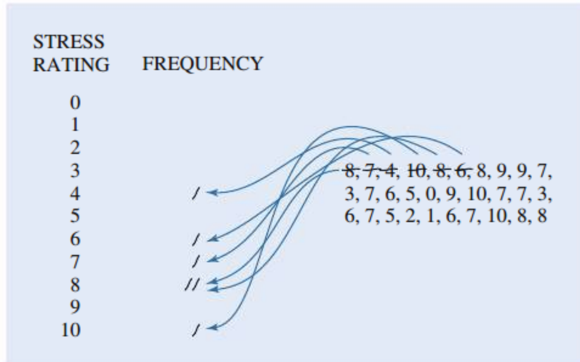
Stress Rating	Frequency	Percent
0	1	3.3
1	1	3.3
2	1	3.3
3	2	6.7
4	1	3.3
5	2	6.7
6	4	13.3
7	7	23.3
8	5	16.7
9	3	10.0
10	3	10.0

SOURCE: Data based on Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112

[1]In addition, some frequency tables include, for each value, the total number of scores with that value and all values preceding it. These are called *cumulative frequencies* because they tell how many scores are accumulated up to this point on the table. If percentages are used, cumulative percentages also may be included. Cumulative percentages give, for each value, the percentage of scores up to and including that value. The cumulative percentage for any given value (or for a score having that value) is also called a *percentile*. Cumulative frequencies and cumulative percentages help you see where a particular score falls in the overall group of scores.

Figure 1–1 A frequency table being prepared by hand for the stress rating scores.

SOURCE: Data based on Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112



❹ **Figure the percentage of scores for each value.** To do this, take the frequency for that value, divide it by the total number of scores, and multiply by 100. The resulting frequency table for the stress rating example is shown in Table 1–3. You will usually need to round off the percentage. We recommend that you round percentages to one decimal place. Note that as a result of the rounding, your percentages will not usually add up to exactly 100% (but the total should be very close).

1.4.3 Frequency Tables for Categorical Variables

The steps we just went over assume you are using numeric variables; however, you can also use a frequency table to show the number of scores in each value (i.e., for each category) of a categorical variable. For example, researchers (Aron, Aron, & Smollan, 1992) asked 208 students to name the closest person in their lives. As shown in Table 1–4, 33 students selected a family member, 76 a nonromantic friend, 92 a romantic partner, and 7 some other person. As you can see in Table 1–4, the values listed on the left-hand side of the frequency table are the values (the categories) of the variable.

1.4.4 An Example about Student Social Interactions

McLaughlin-Volpe, Aron, and Reis (2001) asked 94 first- and second-year university students to keep a diary of their social interactions for a week during the regular semester. Each time a student had a social interaction lasting 10 minutes or longer, the student would fill out a card. The card included questions about who the other people were in the interaction and about various aspects of the conversation. Excluding family and work situations, the number of social interactions of 10 minutes or longer over a week for these 94 students were as follows:

48, 15, 33, 3, 21, 19, 17, 16, 44, 25, 30, 3, 5, 9, 35, 32, 26, 13, 14, 14, 47, 47, 29, 18, 11, 5, 19, 24, 17, 6, 25, 8, 18, 29, 1, 18, 22, 3, 22, 29, 2, 6, 10, 29, 10, 21, 38, 41, 16, 17, 8, 40, 8, 10, 18, 7, 4, 4, 8, 11, 3, 23, 10, 19, 21, 13, 12, 10, 4, 17, 11, 21, 9, 8, 7, 5, 3, 22, 14, 25, 4, 11, 10, 18, 1, 28, 27, 19, 24, 35, 9, 30, 8, 26.

Now, let's follow our four steps for making a frequency table.

❶ **Make a list of each possible value down the left edge, starting from the lowest and ending with the highest.** The lowest possible number of interactions is 0. In this study, the highest number of interactions could be any number. However, the highest actual number in this group

Table 1–4 Frequency Table for a Categorical Variable

Closest Person	Frequency	Percent
Family member	33	15.9
Nonromantic friend	76	36.5
Romantic partner	92	44.2
Other	7	3.4

SOURCE: Data from Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63, 596–612.

Figure 1–2 Making a frequency table of students' social interactions over a week.

SOURCE: Data from McLaughlin-Volpe, T., Aron, A, & Reis, H. T. (2001, February). Closeness during interethnic social interactions and prejudice: A diary study. Paper presented at the Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX.

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

was 48, so we can use 48 as the highest value. Thus, the first step is to list these values. (It might be good to use several columns so that you can have all the scores on a single page or screen.)

❷ **Go one by one through the scores, making a mark for each next to its value on your list.** Figure 1–2 shows the results of this step.

❸ **Make a table showing how many times each value on your list was used.** Table 1–5 is the result.

❹ **Figure the percentage of scores for each value.** We have *not* done so in this example because with so many categories, it would not help much for seeing the pattern of scores. However, if you want to check your understanding of this step, the first five percentages would be 0.0%, 2.1%, 1.1%, 5.3%, and 4.3%. (These are the percentages for frequencies of 0, 2, 1, 5, and 4, rounded to one decimal place.)

Tip for Success Be sure to check your work by adding the frequencies for all of the scores. This sum should equal the total number of scores you started with.

Table 1–5 Frequency Table for Number of Social Interactions During a Week

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

SOURCE: Data from McLaughlin-Volpe, T., Aron, A, & Reis, H. T. (2001, February). Closeness during interethnic social interactions and prejudice: A diary study. Paper presented at the Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX.

1.4.5 Grouped Frequency Tables

Sometimes there are so many possible values that a frequency table is too awkward to give a simple picture of the scores. The last example was a bit like that, wasn't it? The solution is to make groupings of values that include all values in a certain range. For example, consider our stress example. Instead of having a separate frequency figure for the group of students who rated their stress as 8 and another for those who rated it as 9, you could have a combined category of 8 and 9. This combined category is a range of values that includes these two values. A combined category like this is called an **interval**. This particular interval of 8 and 9 has a frequency of 8 (the sum of the 5 scores with a value of 8 and the 3 scores with a value of 9).

A frequency table that uses intervals is called a *grouped frequency table*. Table 1–6 is a grouped frequency table for the stress ratings example. (However, in this example, the full frequency table has only 11 different values. Thus, a grouped frequency table was not really necessary.) Table 1–7 is a grouped frequency table showing the number of social interactions of 94 students over a week.

A grouped frequency table can make information even more directly understandable than an ordinary frequency table can. Of course, the greater understandability of a grouped frequency table is at a cost. You lose information about the breakdown of frequencies within each interval.

When setting up a grouped frequency table, it makes a big difference how many intervals you use. There are guidelines to help researchers with this, but in practice it is done automatically by the researcher's computer (see this chapter's "Using SPSS" section for instructions on how to create frequency tables using statistical software). Thus, we will not focus on it in this book. (Our emphasis in this section is to help you understand a grouped frequency table when you see it in a research article or in the media.) However, should you have to make a grouped frequency table on your own, the key is to experiment with the interval size until you come up with one that is a round number (such as 2, 3, 5, or 10) and that creates about 5–15 intervals. Then, when actually setting up the table, be sure you set the *start of each interval* to be a multiple of the interval size and the top end of each interval to the number just below the start of the next interval. For example, Table 1–6 uses six intervals with an interval size of 2. The intervals are 0–1, 2–3, 4–5, 6–7, 8–9, and 10–11. Note that each interval *starts* with a multiple of 2 (0, 2, 4, 6, 8, 10) and the top end of each interval (1, 3, 5, 7, 9) is the number just below the start of the next interval (2, 4, 6, 8, 10). Table 1–7 uses 10 intervals with an interval size of 5. The intervals are 0–4, 5–9, 10–14, 15–19, and so on, with a final interval of 45–49. Note that each

Table 1–6 Grouped Frequency Table for Stress Ratings

Stress Rating Interval	Frequency	Percent
0–1	2	6.7
2–3	3	10.0
4–5	3	10.0
6–7	11	36.7
8–9	8	26.7
10–11	3	10.0

SOURCE: Data based on Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112

Table 1–7 Grouped Frequency Table for Social Interactions During a Week for 94 College Students

Interval	Frequency	Percent
0–4	12	12.8
5–9	16	17.0
10–14	16	17.0
15–19	16	17.0
20–24	10	10.6
25–29	11	11.7
30–34	4	4.3
35–39	3	3.2
40–44	3	3.2
45–49	3	3.2

SOURCE: Data from McLaughlin-Volpe, T., Aron, A, & Reis, H. T. (2001, February). Closeness during interethnic social interactions and prejudice: A diary study. Paper presented at the Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX.

interval *starts* with a multiple of 5 (0, 5, 10, 15, etc.) and that the top end of each interval (4, 9, 14, 19, etc.) is the number just below the start of the next interval (5, 10, 15, 20, etc.).

II. How are you doing?

Answers can be found at the end of this chapter.

1. What is a frequency table?
2. Why would a researcher want to make a frequency table?
3. Make a frequency table for the following scores: 5, 7, 4, 5, 6, 5, 4.
4. What values are grouped in a frequency table group?

Histograms

1.5 Create histograms from frequency tables

A graph is another good way to make a large group of scores easy to understand. A picture may be worth a thousand words but may also be worth a thousand numbers. A straightforward approach is to make a graph of the frequency table. One kind of graph of the information in a frequency table is a kind of bar chart called a *histogram*. In a histogram, the height of each bar is the frequency of each value in the frequency table.

1.5.1 How to Make a Histogram

There are four steps in making a histogram.

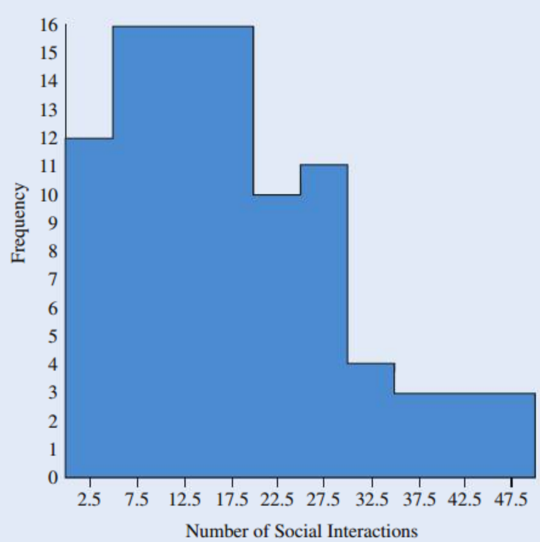
❶ **Make a frequency table (or grouped frequency table).**

❷ **Put the values along the bottom.** The numbers should go from left to right, from lowest to highest. If you are making a histogram from a grouped frequency table, the values you put along the bottom are the interval midpoints. The midpoint of an interval is halfway between the start of that interval and the start of the next highest interval. So, in Figure 1–3, the midpoint for the 0–4 interval is 2.5, because 2.5 is halfway between 0 (the start of the interval) and 5 (the start of the next highest interval). For the 5–9 interval, the midpoint is 7.5, because 7.5 is halfway between 5 (the start of the interval) and 10 (the start of the next highest interval). Do this for each interval.

Figure 1–3 Histogram for the number of social interactions during a week for 94 college students, based on grouped frequencies.

SOURCE: Data from McLaughlin-Volpe, T., Aron, A, & Reis, H. T. (2001, February). Closeness during interethnic social interactions and prejudice: A diary study. Paper presented at the Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX.

INTERVAL	FREQUENCY
0–4	12
5–9	16
10–14	16
15–19	16
20–24	10
25–29	11
30–34	4
35–39	3
40–44	3
45–49	3



❸ **Make a scale of frequencies along the left edge.** The scale should go from 0 at the bottom to the highest frequency for any value.

❹ **Make a bar for each value.** The height of each bar is the frequency of the value over which it is placed. For each bar, make sure that the middle of the bar is above its value.

1.5.2 An Example of Histograms for the Social Interaction Study

Ordinarily, in a histogram all the bars are put next to each other with no space in between, so that it looks a bit like a city skyline. Figure 1–4 shows two histograms based on the stress ratings example, one based on the ordinary frequency table and one based on the grouped frequency table.

Figure 1–4 Four steps in making a histogram based on (a) a frequency table and (b) a grouped frequency table for the stress ratings example. ❶ Make a frequency table. ❷ Put the values along the bottom. ❸ Make a scale of frequencies along the left edge. ❹ Make a bar for each value.

SOURCE: Data based on Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112

(a) Frequency Table

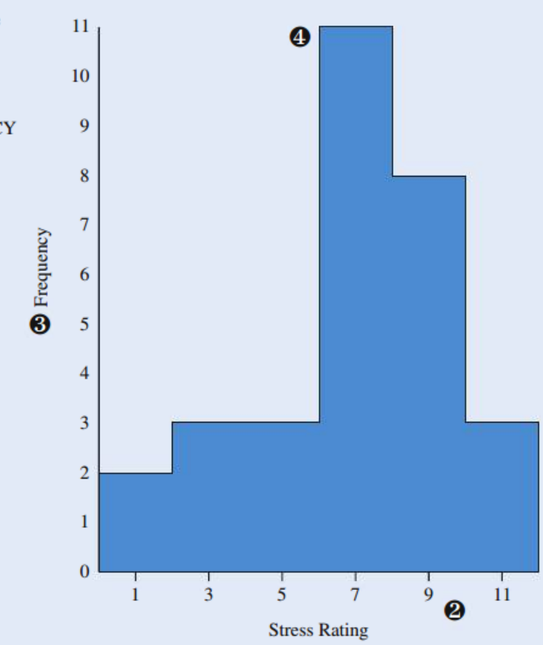
❶

STRESS RATING	FREQUENCY
0	1
1	1
2	1
3	2
4	1
5	2
6	4
7	7
8	5
9	3
10	3

(b) Grouped Frequency Table

❶

STRESS RATING INTERVAL	FREQUENCY
0–1	2
2–3	3
4–5	3
6–7	11
8–9	8
10–11	3



Tip for Success You will probably find it easier to make histograms by hand if you use graph paper.

1.5.3 Bar Graphs and Categorical Variables

When you have a categorical variable, the histogram is called a bar graph. Since the values of a categorical variable are not in any particular order, you leave a space in between the bars. Figure 1–5 shows a bar graph based on the frequency table in Table 1–4.

Tip for Success In some research articles, you may see a histogram of a numeric variable with a space in between the bars. We recommend, however, that you only leave a space in between the bars when you are creating a bar graph of a categorical variable.

III. How are you doing?

Answers can be found at the end of this chapter.

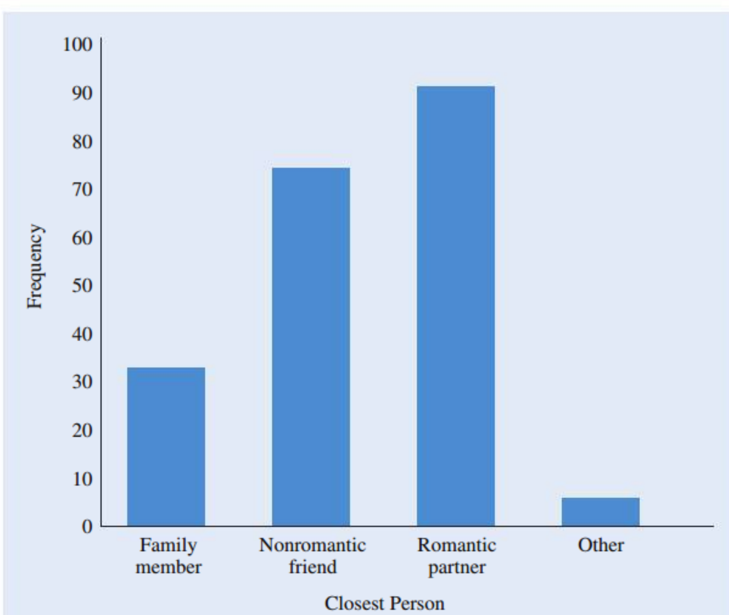
1. Why do researchers make histograms?
2. When making a histogram from a frequency table, (a) what goes along the bottom, (b) what goes along the left edge, and (c) what goes above each value?
3. Make a histogram based on the frequency table below.

Value	Frequency
1	3
2	4
3	8
4	5
5	2

4. How is a histogram based on a categorical variable different from one based on a number equal-interval variable?

Figure 1–5 Bar graph for the closest person in life for 208 students (see Table 1–4).

SOURCE: Data from Aron, A., Paris, M., & Aron, E. N. (1995). Falling in love: Prospective studies of self-concept change. *Journal of Personality and Social Psychology*, 69, 1102–1112



Shapes of Frequency Distributions

1.6 Identify the shape of a distribution of data

A *frequency distribution* shows the pattern of frequencies over the various values. A frequency table or histogram describes a frequency distribution because each shows the pattern or shape of how the frequencies are spread out, or "distributed." Behavioral and social scientists also describe this shape in words. Describing the shape of a distribution is important both in the descriptive statistics we focus on in this chapter and also in the inferential statistics you will learn in later chapters.

1.6.1 Unimodal and Bimodal Frequency Distributions

One important aspect of a distribution's shape is whether it has only one main high point: one high "tower" in the histogram. For example, in the stress ratings study, the most frequent score is a 7, giving a histogram with only one very high area. This is called a *unimodal distribution*. If a distribution has two fairly equal high points, it is called a *bimodal distribution*. Any distribution with two or more high points is called a *multimodal distribution*. Finally, if all the values have about the same frequency, it is called a *rectangular distribution*. Figure 1–6 shows examples of these frequency distribution shapes.

As you will see, the graphs in Figure 1–6 are not histograms, but special line graphs called frequency polygons, which are another way to graph a frequency table. In a *frequency polygon*, the lines move from point to point. The height of each point shows the number of scores with that value. This creates a mountain peak skyline.

The information we collect in behavioral and social science research is usually approximately unimodal. Bimodal and other multimodal distributions occasionally turn up. A bimodal example is the distribution of the ages of people in a toddler's play area in a park, who are mostly either toddlers with ages of around 2–4 years or caretakers with ages of 20–45 years or so (with few infants, a few people aged 5–19 years, and a few above 45). Thus, if you make a frequency distribution of these ages, the large frequencies would be at the values for toddler ages (2–4 years) and for higher ages (20–40 years or so). An example of a rectangular distribution is the number of children at each grade level attending an elementary school: there is about the same number in first grade, second grade, and so on. Figure 1–7 shows these examples.

1.6.2 Symmetrical and Skewed Distributions

Look again at the histograms of the stress rating example (Figure 1–4). The distribution is lopsided, with more scores near the high end. This is somewhat unusual. Most things we measure in the behavioral and social sciences have about equal numbers on both sides of the middle. That is, most of the time, the scores follow an approximately *symmetrical distribution* (if you fold the graph of a symmetrical distribution in half sideways, the two halves look the same).

Figure 1–6 Examples of (a) unimodal, (b) approximately bimodal, and (c) approximately rectangular frequency polygons.

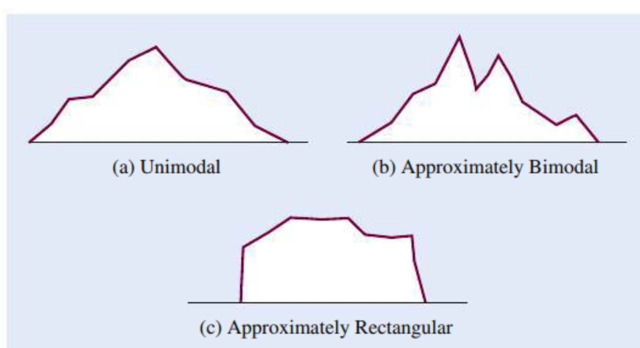
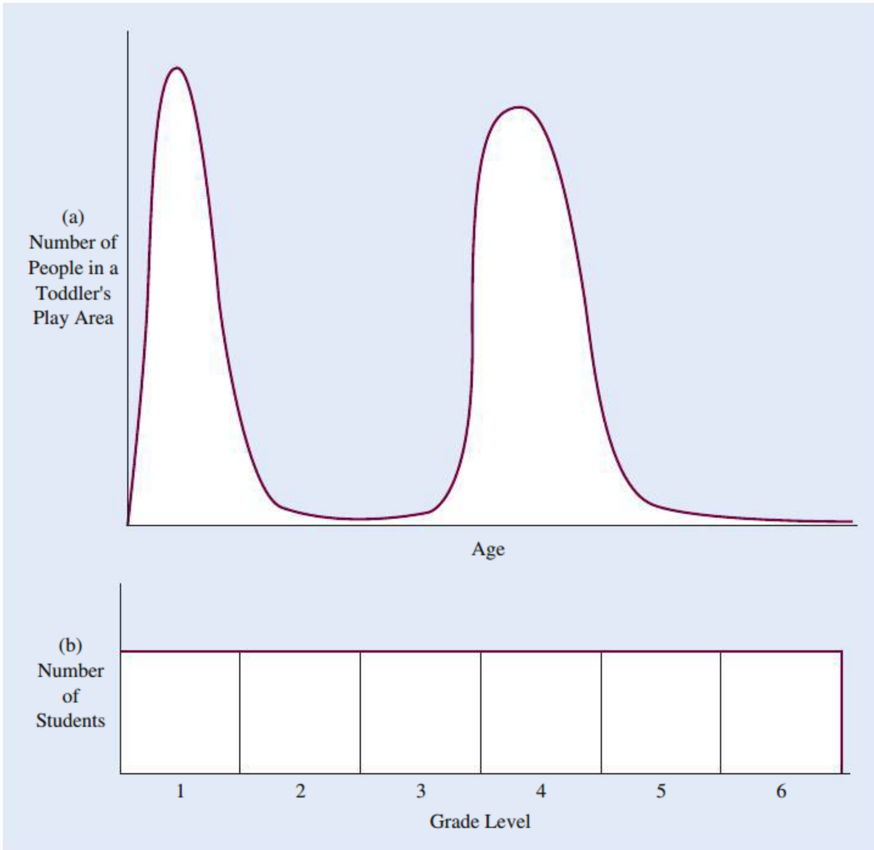


Figure 1–7 Fictional examples of distributions that are not unimodal: (a) A bimodal distribution showing the possible frequencies for people of different ages in a toddler's play area. (b) A rectangular distribution showing the possible frequencies of students at different grade levels in an elementary school.



A distribution that is clearly not symmetrical is called a *skewed distribution*. The stress ratings distribution is an example. A skewed distribution has one side that is long and spread out, somewhat like a tail. The side with *fewer* scores (the side that looks like a tail) describes the direction of the skew. Thus a distribution with fewer scores left of the peak, like our stress ratings example, is *skewed to the left*. The other example we have examined in this chapter, the distributions of students' numbers of interactions in a week, is *skewed to the right*. Figure 1–8 shows examples of approximately symmetrical and skewed distributions.

Tip for Success You may be interested to know that the word *skew* comes from the French *queue*, which means "line or tail." This should help you remember that the direction of the skew (to the left or right) is the side that has the long line or tail.

Figure 1–8 Examples of frequency polygons of distributions that are (a) approximately symmetrical, (b) skewed to the right (positively skewed), and (c) skewed to the left (negatively skewed).