Cover image: iStock.com/AF-studio

ISBN 978-1-5443-9800-6

SAGE college publishing

BABBIE • WAGNER • ZAINO

Adventures in SOCIAL RESEARCH

EDITION 11

EARL BABBIE • WILLIAM E. WAGNER, III • JEANNE ZAINO

Adventures in SOCIAL RESEARCH

EDITION 11

DATA ANALYSIS USING IBM® SPSS® STATISTICS

SAGE

ADVENTURES in SOCIAL RESEARCH

Eleventh Edition

To our students:
Past, present, and future—
We challenge each other and profit from it.

ADVENTURES in SOCIAL RESEARCH

Data Analysis Using IBM® SPSS® Statistics

Eleventh Edition

Earl Babbie
Chapman University

William E. Wagner, III
California State University, Channel Islands

Jeanne Zaino
Iona College

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Printed in the United States of America

ISBN: 978-1-5443-9800-6

This book is printed on acid-free paper.

Acquisitions Editor: Helen Salmon
Product Associate Ivey Mellem
Production Editor: Astha Jaiswal
Copy Editor: Diane DiMura & TC
Typesetter: C&M Digitals (P) Ltd.
Cover Designer: Scott Van Atta
Marketing Manager: Victoria Velasquez

22 23 24 25 26 10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

DETAILED CONTENTS

PREFACE

This workbook is offered to you with a number of aims in mind. To begin, we want to provide students with a practical and hands-on introduction to the logic of social science research, particularly survey research. Moreover, we want to give students an accessible book that guides them, step by step, through the process of data analysis using current General Social Survey (GSS) data and the latest versions of IBM® SPSS® Statistics* for either Windows or Macintosh computers (as well as SPSS utilized through virtual computing labs). Most importantly, we want to involve students directly in the practice of social research, allow them to experience the excitement and wonder of this enterprise, and inspire them to pursue their own adventures in social research.

As we pursue these goals, however, there are a number of agendas in the background of this book. For example, students who complete the book will have learned a very useful, employable skill. Increasingly, job applicants are asked about their facility with various computer programs: word processing, spreadsheets, and data analysis. As of this writing, SPSS Statistics is still clearly the most popular professional program available for social science data analysis—hence, our choice of it as a vehicle for teaching social research.

WHAT'S NEW IN THE 11TH EDITION OF *ADVENTURES IN SOCIAL RESEARCH*?

All of the demonstrations, examples, and lab problems throughout the entire book have been revised or replaced to reflect the latest General Social Survey (GSS) data available at the time of writing (2018 GSS). In addition, the screenshots and output used throughout the book have been produced on the latest versions of SPSS available at the time of writing (v.27 and v.28). These changes can be seen in almost every chapter and most of the chapters are different in substantial ways given that the results are all different; in fact, for some inquiries the difference is not small, but can yield a finding in a new direction altogether!

Since the software and the data we use are entirely new in this edition, we will also be providing newly created screencast videos to accompany many of the chapter lab exercises.

We have also taken care to integrate inclusive language and examples throughout this book. Sometimes the data may show that the opinions of a large share of the U.S. population may not be sensitive to principles of diversity, but the way we approach that discussion is always from a place of respect and inclusiveness.

A FOCUS ON DEVELOPING PROFESSIONAL AND INTELLECTUAL SKILLS

What sets this book apart from others that teach SPSS Statistics or similar programs is that we present this particular skill within the context of social research as a logical enterprise.

Thus, in addition to learning to use SPSS Statistics, students are learning the intellectual "skills" of conceptualization, measurement, and association. Even though those who know only SPSS Statistics can assist in data analysis, our intention is that students will also be able to think for themselves, mapping out analytic paths into the understanding of social data. As they polish these intellectual

*SPSS is a registered trademark of International Business Machines Corporation.

skills, they should be able to progress to higher levels of research and to the administration of research enterprises.

More generally, we aim to educate students who *will use* computers rather than *be used by* them. In our experience, when students first confront computers in school, they tend to fall into two groups: those who recognize computers as powerful instruments for pursuing their goals in life, or at least as the grandest of toys, and those who are intimidated by computers and seek the earliest possible refuge from them. Our intention is to reveal the former possibility to students and to coax them into a productive relationship with computers.

Educators are being challenged increasingly to demonstrate the practical value of instruction, no less in the social sciences than in other fields. Too often, an overreaction to this demand results in superficial vocational courses that offer no intellectual meaning or courses hastily contrived as a home for current buzzwords, whose popularity is often short-lived. We are excited to be able to offer an educational experience that is genuinely practical for students and that also represents an intellectual adventure.

Those who have taught research methods or statistics courses typically find themselves with a daunting task: to ignite their often involuntary students with the fire of enthusiasm they feel for the detective work of social research at its best. In this book, we seek to engage students' curiosity by setting them about the task of understanding issues that are already points of interest for them: topics such as abortion, religion, politics, poverty, gender roles, sexual attitudes, mass media, gun control, child-rearing, racial prejudice, and others. For many of our readers, we imagine that *mathematical analysis* still smacks of trains leaving Point A and Point B at different speeds and so on. Now, they are going to learn that some facility with the logic and mathematics of social research can let them focus the light of understanding on the dark turbulence of opinion and hysteria.

We do not tell students about opinions on abortion as much as we show them how to find out for themselves. We think that will get students to Point C ahead of either of the trains.

A FOCUS ON ACTIVE AND COLLABORATIVE LEARNING

As we are teaching students to learn for themselves, this book offers a good example of what educators have taken to calling *active learning*. We have set up all of our demonstrations so that students should be executing the same SPSS Statistics operations we are discussing at any given point. Although we may give them the "answers" to assure them that they are on the right track, we leave them on their own often enough to require that they do the work rather than simply read about it.

Finally, the culture of personal computers has been one of *collaborative learning* from its very beginning. More than people in any other field of activity, perhaps, computer users have always delighted in sharing what they know with others. There is probably no better context within which to ask for help: Those who know the answer are quick to respond, and those who do not often turn their attention to finding an answer, delighting in the challenge.

Because this book is self-contained, even introductory students can walk through the chapters and exercises on their own, without outside assistance. We imagine, however, that students will want to work together as they progress through this book. That has been our experience in student testing and in earlier courses we have taught involving computers. We suggest that you encourage cooperation among students; we are certain that they will learn more that way and will enjoy the course more. In fact, those who are initially intimidated by computers should especially be encouraged to find buddies with whom to work.

INTENDED FOR STUDENTS IN VARIOUS SOCIAL SCIENCE DISCIPLINES

This book is intended for use in any social science course that either introduces or focuses exclusively on social research methods, social statistics, data analysis, or survey research. It can be easily combined

with or used as a supplement to most standard social science textbooks, including but not limited to those in fields as varied as communication science, criminal justice, health studies, political science, public policy, social work, and sociology.

As far as possible, we have designed this book to be "self-writing" and "open-ended" to ensure it is relevant to students with varying interests across numerous disciplines. Throughout the text, we encourage students to focus on issues and questions relevant to their particular areas of interest. After walking through the demonstrations that introduce the fundamentals of the data analysis process, students are given a chance to apply what they have learned. In many of the lab exercises, students are encouraged to design their own hypotheses, choose their own variables, and interpret the results. Moreover, we encourage instructors to apply the principles, techniques, and methods discussed to other data sets relevant to their fields.

INTENDED FOR BOTH BEGINNING AND MORE ADVANCED STUDENTS

We have designed and structured this book to support students at a variety of levels. This includes both those students who are taking their first course in social research and more advanced students (including graduate students) who either want to hone their social research, statistical, and data analysis skills or merely want to become acquainted or reacquainted with the latest versions of SPSS Statistics for Windows or Macintosh. More advanced students who come at this book full speed may choose either to work through the text from beginning to end or to skip around and focus on particular chapters and sections.

It is important to note, however, that because this book is "self-contained" and guides the student analyst step by step through the demonstrations and exercises, no previous experience with social research, statistics, computers, Windows, Macintosh, or SPSS Statistics is required. Those who have never taken a research methods, statistics, or computer-based course will find that they can easily progress through this book.

THE BOOK AND THE FREE COMPANION WEBSITE: WHAT IS INCLUDED?

The book and *Adventures in Social Research* companion website provide the instructions and data needed to introduce students to social science data analysis. Most college and university computing services make SPSS Statistics software available to students. The companion website includes the data sets that are used throughout the book: DEMO.SAV and EXER.SAV, which can be analyzed by most versions of SPSS Statistics on Windows or Macintosh computers. As you will see, although we have provided for analysis in some depth in a few instances, the variables cover a fairly broad terrain. In addition to working their way through the demonstrations and exercises presented in the book, students will be able to find original lines of inquiry that grow out of their own interests and insights.

This book will illustrate the use of SPSS Statistics, using version 27.0 for Windows and Macintosh. While the text focuses specifically on the current version of SPSS Statistics, it can also be easily used with recent versions. Regardless of the version you are using, throughout the text, we will refer to the program simply as "SPSS Statistics," which is identical to PASW Statistics. (Versions 17 and 18 were known as PASW as part of a branding shift, before SPSS was acquired by IBM in 2009 and the name reverted to SPSS.) SPSS Statistics comes with extensive help screens. They are almost like having a coach built into your computer! Begin with the menu farthest to the right.

You can click **Help** to see the options available to you. "Topics" will usually be your most useful choice. This will give you three options: "Contents," "Index," and "Search." "Contents" and "Index" present you with two ways of zeroing in on the topic of interest to you. "Search" will search for the specific terms or keywords you indicate. You should experiment with these options to discover what works best for you.

ORGANIZATION AND CONTENT

The chapters are arranged in an order that roughly parallels the organization of most introductory social science research methods texts. Part I (Chapters 1–3) includes an overview of the essentials of social research and description of the 2018 GSS from which comes the data file you will be working (ADVENTURES.SAV). Parts II, III, and IV (Chapters 4–20) introduce readers to SPSS Statistics and data analysis, beginning with univariate analysis, then bivariate analysis, and finally multivariate analysis, respectively. Part V (Chapters 21–22) focuses on primary research and additional avenues for secondary research.

Part I includes three chapters that help prepare students for social research using the General Social Survey. Our goal in these chapters is to give students an introduction to some of the fundamental elements of social scientific research, particularly those they will encounter later in the text.

Chapter 1 discusses the main purposes of the text and introduces students to some of the historical background that lies behind computerized social research, data analysis, and statistical software packages. It also introduces readers to the process of social research. Chapter 2 continues this focus by examining the logic of important aspects of social research, including theory, research, and measurement. Chapter 3 is designed to introduce readers to the GSS and the data sets included with this book.

Part II is designed to help students get started using SPSS Statistics and to give them practice with the basics of univariate analysis. Chapter 4 introduces students to SPSS Statistics by guiding them through the steps involved in launching the program, opening their data sets, and exploring the variables contained on the disk that accompanies this book.

The rest of this section (Chapters 5–9) is devoted to data analysis. In Chapter 5, we introduce frequency distributions, descriptive statistics, and saving and printing data. Chapter 6 focuses on the graphic presentation of univariate data by covering the commands for creating bar and pie charts, line graphs, and histograms. In Chapters 7 and 8, students are given a chance to practice recoding and creating composite variables. The final chapter in this section (Chapter 9) allows students to strike out on their own and apply the methods and techniques they have learned in this section to other topics.

Part III focuses primarily on bivariate analyses. In Chapters 10 through 12, we limit our discussion to the analysis of percentage tables. In Chapters 13, 14, and 15, we introduce other methods for examining the extent to which two variables are related to each other.

Chapter 13 focuses on some common measures of association for nominal and ordinal variables, including lambda and gamma. Chapter 14 focuses on correlation and regression analysis, by exploring Pearson's r, r^2, and OLS (ordinary least squares) regression. Chapter 15 introduces tests of statistical significance, such as chi-square, t tests, and ANOVA. In Chapter 16, students are once again given a chance to apply the bivariate techniques and methods discussed in Part III to other topics and issues.

Our discussion of data analysis concludes in Part IV with a discussion of multivariate analyses. Chapter 17 focuses primarily on multiple causation. Chapter 18 picks up on some of the loose threads of our bivariate analyses and pursues them further, while Chapter 19 guides students through the steps involved in creating composite measures to predict opinions. Finally, in Chapter 20, students are given a chance to apply the methods and techniques discussed in Part IV to other topics and issues.

The final section comprises two chapters that explore some further opportunities for social research. Because students often express an interest in collecting their own data, Chapter 21 focuses on primary research. We introduce students to the steps involved in designing and administering a survey, defining and entering data in SPSS Statistics, and writing a research report. This chapter is supplemented by several articles on the *Adventures in Social Research* companion website that are intended to give students additional information regarding preparing a research proposal, designing and administering a survey, constructing a sample questionnaire, and writing a research report. Chapter 22 suggests other avenues for pursuing secondary social research by focusing on the unabridged GSS, additional data sources, and other statistical software packages that students may find useful.

The text contains a codebook (Appendix A) that describes all of the variables in the data files used in this book. The full GSS Codebook is included on the companion website, along with the full GSS file for 2018. A sample questionnaire designed for student use (with instructions in Chapter 21) is provided in Appendix B. We have also updated and expanded the references, index, and glossary.

The *Adventures in Social Research* companion website contains not only our data sets (Adventures.SAV, GSS2018.SAV, and Students.SAV) but also articles relevant to the discussion of primary research in Chapter 21, a comprehensive list of all of the SPSS Statistics Commands introduced in the text, and some recommended readings that relate to topics and issues covered in the text.

STRUCTURE OF EACH CHAPTER

Chapters include explanations of basic research principles and techniques; specific instructions regarding how to use SPSS Statistics; and demonstrations, writing boxes, a brief conclusion, a list of main points, key terms, SPSS Statistics commands introduced in the chapter, review questions, and SPSS Statistics lab exercises. Students are expected to follow along with the demonstrations in the body of each chapter. They are aided in this process by both the text, which walks them, step by step, through the process of data analysis, and screen captures, which help them understand what they should be seeing on their own monitor.

In an effort to stress the importance of describing research findings in prose, most chapters include writing boxes, which give readers an example of how a professional social scientist might describe the findings being discussed. The review questions at the end of each chapter are designed to test the students' knowledge of the material presented in the text. Because they do not require SPSS Statistics, these questions can be assigned as either class work or homework assignments.

In the SPSS Statistics lab exercises, students are given a chance to apply what they learned in the explanatory sections and demonstrations. These exercises generally follow a fill-in-the-blank format for presenting, analyzing, and summarizing results. Instructors may wish to assign these exercises as lab assignments to be completed either in the lab or as homework, provided students have access to SPSS Statistics, either the full or the student version, on Windows or Macintosh.

Although the book is designed to guide students through the process of computerized data analysis from beginning to end, we encourage instructors—and particularly more advanced students—to skip around and focus on chapters and sections that are of interest to them. We designed the book with the understanding that students at various levels may find different demonstrations, techniques, discussions, and methods of varying interest. Consequently, all of the chapters are self-contained, and both students and instructors should feel comfortable picking and choosing among topics, issues, and material of particular interest to them.

Instructors and students who choose to take this approach may want to refer to the table of contents, introductions to each part, chapter conclusions, and summaries of main points to get a better sense of which sections and chapters they want to focus on.

SOFTWARE SUPPORT AND SERVICE

If you or your students should run into any problems using this package, there are several sources of support that should serve your needs. Frequently, college and university computing centers have student assistants who are very helpful to new computer users. In fact, most academic computing centers employ a user services coordinator who can help faculty plan student use of the school's computers and provide aid when problems arise.

One source of SPSS Statistics assistance available via the internet is the product's Customer Resource page at https://www.ibm.com/analytics/us/en/technology/spss. (SPSS has been an IBM company since

2009.) In addition to providing answers to frequently asked questions, it offers a variety of tips and white papers on important issues in data analysis. Registered users may submit specific questions to consultants via an online form at the Business Analytics Client Center. SPSS Statistics requests that a legitimate license or serial number be submitted with questions in order to receive a response. Be forewarned that SPSS cannot give assistance with pedagogical or substantive problems. It has been our experience that the best help comes from local resources.

ACKNOWLEDGMENTS

In conclusion, we would like to acknowledge a number of people who have been instrumental in making this book a reality. First on the list is Jerry Westby of SAGE/Pine Forge Press. Our thanks also go to former Pine Forge Press editor Steve Rutter and the many others at SAGE/Pine Forge Press who aided us along the way, particularly Claudia Hoffman, Kristin Snow, and Vonessa Vondera for their work on the fifth edition and Denise Santoyo, Cheryl Rivard, and Kim Suarez for their work on the sixth edition. Special thanks go also to Megan Granger and Nichole O'Grady for their work on the seventh edition, and again to Megan Granger for her work on the eighth edition. Paula Fleming was our copyeditor for the ninth edition, Jared Leighton for the tenth edition, and Diane DiMura and TC for the eleventh edition.

We would also like to thank the many reviewers who helped us along the way: Monika Ardelt, University of Florida; Karen August, Humboldt State University; Dhruba J. Bora, Wheeling Jesuit University; Tom Buchanan, University of Tennessee at Chattanooga; Campbell J. Bullock, California State University, Leia K. Cain, University of Tennessee; Stanislaus; Gregg Lee Carter, Bryant University; James Cassell, Indiana State University; Christina DeJong, Michigan State University; Xiaogang Deng, University of Massachusetts, Boston; Jamie Dolan, Carroll College; Lucas Espinoza, University of Texas Rio Grande Valley; Stephen J. Farnsworth, Mary Washington College; Ellen Granberg, Clemson University; Jessica Greene, University of Oregon; John R. Hagen, Health Strategies Inc.; Jade Hsuan Huang, Hawaii Pacific University; Ting Jiang, Metropolitan State University of Denver; Steven L. Jones, University of Virginia; Quentin Kidd, Christopher Newport University; Wanda Kosinski, Ramapo College of New Jersey; Mary Jane Kuffner Hirt, Indiana University of Pennsylvania; Peter Allen Lee, San Jose State University; Kim Lloyd, Washington State University; William H. Lockhart, University of Virginia; Karyn D. McKinney, Penn State University, Altoona College; Maureen McLeod, Sage Colleges; Fred E. Markowitz, Northern Illinois University; Josh Meisel, Humboldt State University; Philip Meyer, University of North Carolina; Amanda Miller, University of Indianapolis; David B. Miller, Case Western Reserve University; Ojmarrh Mitchell, University of Nevada–Las Vegas; Melanie Moore, University of Northern Colorado; Thomas O'Rourke, University of Illinois at Urbana-Champaign; Graham Ousey, University of Delaware; Linda Owens, University of Illinois; Hyung Sam Park, University of Central Florida; Boyka Stefanova, University of Texas, San Antonio; Sue Strickler, Eastern New Mexico University; Mangala Subramaniam, Purdue University; Jennifer L. S. Teller, Kent State University; Lanny Thomson, University of Puerto Rico; John Tinker, California State University–Fresno; Madine VanderPlaat, Saint Mary's University; Kristie Vise, Northern Kentucky University; Roland Wagner, San Jose State University; Joshua J. Weikert, West Chester University; David Zeller, University of South Florida; Assata Zerai, Syracuse University; and Thomas G. Zullo, University of Pittsburgh.

We reserve our final acknowledgment for our students, to whom this book is dedicated. We recognize that we have often asked them to think and do things that they sometimes felt were beyond their abilities. We have admired their courage for trying anyway, and we have shared in their growth.

ABOUT THE AUTHORS

Earl Babbie was born in Detroit, Michigan, in 1938, but his family chose to return to Vermont three months later, and he grew up there and in New Hampshire. In 1956, he set off for Harvard Yard, where he spent the next four years learning more than he initially planned. After three years with the U.S. Marine Corps, mostly in Asia, he began graduate studies at the University of California, Berkeley. He received his PhD from Berkeley in 1969. He taught sociology at the University of Hawaii from 1968 through 1979, took time off from teaching and research to write full-time for eight years, and then joined the faculty at Chapman University in Southern California in 1987. Although he is the author of several research articles and monographs, he is best known for the many textbooks he has written, which have been widely adopted in colleges throughout the United States and the world. He also has been active in the American Sociological Association (ASA) for 25 years and currently serves on the ASA's executive committee. He is also past president of both the Pacific Sociological Association and the California Sociological Association.

William E. Wagner, III is professor of health sciences and sociology at California State University, Channel Islands, and also the executive director of the Social Science Research and Instructional Council of the California State University System (SSRIC). He has served as a member of the faculty and director of the Institute for Social and Community Research at California State University, Bakersfield, prior to coming to Cal State, Channel Islands. His MA and PhD degrees in sociology are from the University of Illinois, Chicago. In addition, Dr. Wagner earned a master's of public health (MPH) degree from California State University, Northridge. He holds two separate bachelor's degrees, one in mathematics and the other in sociology/anthropology, both from Saint Mary's College of Maryland. His work on topics such as urban sociology, sports, homophobia, academic status, health, and sexual behavior has been published in national and regional scholarly journals. He was elected as vice president of the California Sociological Association for two years and served for seven years as the chair of the Pacific Sociological Association's Committee on the Status of GLBT Persons. Dr. Wagner has also authored *Using IBM SPSS Statistics for Research Methods and Social Science Statistics,* 7th edition (SAGE, 2019), as well as three co-authored books: *The Practice of Survey Research* (2016), with Erin Ruel and Brian Gillespie, *Using and Interpreting Statistics in the Social, Behavioral, and Health Sciences,* with Brian Gillespie, and *A Guide to R for Social & Behavioral Science Statistics*, with Brian Gillespie and Kathleen Hibbert.

Jeanne Zaino, professor of political science, Iona College, earned a bachelor's degree in political science and a master's degree in survey research at the University of Connecticut, Storrs. During that time, she worked as a research assistant at the Roper Center for Public Opinion Research. She went on to earn a master's degree and PhD in political science from the University of Massachusetts, Amherst. She is currently chair of the Political Science Department at Iona College in New Rochelle, New York, where she teaches courses in American government, institutions, research methods, social statistics, public opinion, scope, and methods. She and her husband, Jeff, are the proud parents of two sons, Maxim and Logan.

PART I

PREPARING FOR DATA ANALYSIS

In the opening chapters, we introduce you to computerized data analysis, the logic of social science research, and the data you will be using throughout this text.

In Chapter 1, we begin by looking at how and why social scientists use computers and computer programs. Then, we introduce you to the two main pillars of social research: logic and observation. We will demonstrate how theory—the *logic* component—informs our investigations, makes sense out of our *observations*, and sometimes offers predictions about what we will find. We will also touch on another important aspect of research: the collection and analysis of data, such as those collected in a survey.

In Chapter 2, we delve more deeply into one central component of scientific inquiry: measurement. We look at some of the criteria for measurement quality and start examining the kinds of measurements represented by the data at hand.

In Chapter 3, we introduce you to the data you will be using throughout the text. The data come from the 2018 General Social Survey (GSS). The GSS reflects the attitudes of a representative sample of American adults on a variety of issues, from religion, politics, abortion, and child-rearing to the mass media, sex, law enforcement, and immigration.

INTRODUCTION

The Theory and Practice of Social Research

Social research is the detective work involved in answering big questions. Whereas a conventional detective tries to find out who committed a specific crime, a social researcher looks for the causes of crime in general. The logic of social scientific investigation extends beyond crime to include all aspects of social life, such as careers, marriage and family, voting, health, prejudice, environment, and poverty. In fact, anything that is likely to concern you as an individual is the subject of social science research.

OVERVIEW

The purpose of this book is to lead you through a series of investigative adventures in social research. We can't predict exactly where these adventures will lead because you are going to be the detective. Our purpose is to show you some simple tools (and some truly amazing ones) that you can use in social investigations.

We'll also provide you with a body of data, collected in a national survey, that is so rich you will have the opportunity to undertake investigations that no one else has ever pursued.

If you have access to a computer that uses a recent Microsoft Windows or Mac OS configuration and IBM SPSS Statistics (or the even the older temporarily branded PASW, Predictive Analytics Software), this book and the website associated with it (https://edge.sagepub.com/babbie11e) contain everything you need for a wide range of social investigations.[1] The website is designed specifically for exploring data by way of a fascinating computer game. Instead of fighting off alien attacks or escaping from dank dungeons, you'll be pitting your abilities and imagination against real life, but you'll be looking at a side of life that you may not now be aware of.

This tool is also well designed to aid in the creation of college term papers, and throughout this book, we suggest ways to present the data you discover in the context of a typical term paper in the social sciences. Whereas most students are limited in their term papers to reporting what other investigators have learned about society, you will be able to offer your own insights and discoveries.

Finally, the data set included on the website is being analyzed by professional social scientists today. Moreover, the analytical tools that we've provided for you are as powerful as those used by many professional researchers. Frankly, there's no reason you can't use these materials for original research worthy of publication in a research journal. All it takes is curiosity, imagination, practice, and a healthy obsession with knowing the answers to things. In our experience, what sets professional researchers apart from others is that they have much greater curiosity about the world around them and, in fact, are passionately driven to understand it. They bring a powerful imagination to bear on this task, and they are willing to put in the time necessary for effective investigation.

WHY USE A DATABASE?

As do physical scientists, social scientists use observations of the empirical world to develop and evaluate theories. Much of the social scientist's work involves ascertaining whether logically derived relationships expressed in social theories correspond to empirically observed relationships among social data. For instance, a theory may suggest that Catholics are more opposed to abortion than are non-Catholics, but we don't have scientific evidence until we poll Catholics and non-Catholics and evaluate their differences on abortion. To have confidence in our findings, we must poll a large number of people for their positions on abortion and their religious preferences. And for sure, we probably would feel that any explanation of difference was incomplete until other factors, such as each respondent's gender, age, social class, and so forth, were included. We don't have to stretch our imagination very far to realize that even simple research can soon generate a large mass of data, given the number of cases and variables that we need to provide credible evidence for or against a theory.

Given the sheer number of observations commonly made by social scientists, researchers need to use a database to take into account all relevant data in their research. The full 2018 General Social Survey (GSS) data file we have included with this book, for example, contains more than 1,000 variables for 2,348 individuals. Initial analysis of the data requires that they be sorted, categorized, and recategorized before statistics may be computed for them. With more than 2.3 million data points available for use, we clearly need to use a database to complete a meaningful analysis this semester!

SPSS STATISTICS

Today, the two statistical packages most widely used by social scientists are ***SPSS Statistics*** (originally known as the Statistical Package for the Social Sciences and temporarily branded as ***PASW Statistics*** for version 18; since 2009, IBM SPSS Statistics) and ***SAS (Statistical Analysis System***).[2] We have selected SPSS Statistics for use with this textbook for three reasons. First, early versions of SPSS date back to 1968. The package is well known, and hardly any social scientists who have earned a graduate degree in the past 50 years have not had some contact with SPSS. Second, SPSS takes you through all of the basic issues of using a statistical package. This knowledge will give you a head start if you learn some other package later. Finally, recent versions of SPSS Statistics are suitable for computers running current and recent versions of Microsoft Windows. There are also versions available for Apple Macintosh computers running current and recent versions of Mac OS.

The SPSS Statistics Base is sold as the basic package. Then, if the buyer wishes, it can be upgraded with powerful statistical accessories, all of which are beyond the scope of this book. You can think of it like a car: The base price includes basic features, and then, you can add options or like a great sound system and high-performance tires (or "packages" that include additional groups of equipment and upgrades for the car). Other upgraded software packages include SPSS Professional and—the ultimate model—SPSS Premium.

SPSS also offers two packages specifically designed for students: SPSS Statistics Student Version (offered through version 18 and from versions 22 through 27 but discontinued for version 28) and SPSS Statistics Graduate Pack. Although both versions are also available for use with Windows or Macintosh computers, they differ in terms of their capabilities. Unlike the SPSS Statistics Base system or the SPSS Statistics Graduate Pack, the ***SPSS Statistics Student Version*** is limited to a maximum of 50 variables and a maximum of 1,500 cases; this version cannot be upgraded, and at the time of publishing, we are told that the student version will no longer be produced or supported going forward. The student version has fewer statistical procedures but has most of the procedures that will ever be needed by an undergraduate social science major or a master's-level graduate student, so it is sufficient to use in tandem with this book. If you do happen to be using a student version, you will need to reduce the number of cases to 1,500 and the number of variables to 50. We will make a file available on the study website associated with this book, but please remember that your output will differ slightly from the output presented in the book, mainly because your sample sizes will be different; results may vary slightly, as well, by using only a subsample: 1,500 from over 2,300.

The SPSS Statistics Graduate Pack contains the SPSS Statistics Base system plus two advanced statistical modules. The graduate pack is commonly available at college and university bookstores. You can learn more about IBM SPSS Statistics and the various versions that are available by visiting the company's online store at https://www.ibm.com/analytics/us/en/technology/spss.[3] IBM may change the configurations or offerings of these packages or modules in the future, so be sure to check the website for the latest details.

While having a copy of SPSS Statistics on your personal computer is convenient, you may not need to purchase the software to complete the exercises in this book. Most colleges and universities offer SPSS Statistics to their students through the school's network. More recently, client/server desktop application technology has been adopted at some colleges and universities. With this technology, your school can provide you with client software (e.g., Citrix) and settings to put on your personal Windows or Macintosh computer. You will then be able to run the university's version of SPSS Statistics from its server through your computer's desktop application as though the software were directly installed on your computer. Of course, you will need to remain connected to the Internet, and a high-speed connection is required for this software to be useful. More recently, some colleges and universities have begun to offer "virtual computing labs" where you can control a computer in the "lab" from your own desktop. You can then use any software on that computer. Typically, SPSS is among the software available on lab computers, but if not, be sure to contact your institution's help desk to ask for it to be installed. Check with your school's IT (information technology) office to see if these remote capabilities are available to you. Other universities may offer temporary student licenses for enrolled students, where students are granted a temporary license to install a copy of the base software on their own personal computer without any charge.

SOCIAL RESEARCH: A PRIMER

This book addresses the techniques of social science data analysis. Thus, we're going to be spending most of our time using SPSS Statistics to analyze data and reach conclusions about the people who answered questions in the General Social Survey (GSS), as described in more detail in Chapter 3.

Data analysis, however, doesn't occur in a vacuum. Scientific inquiry is a matter of both observing and reasoning. Consequently, before focusing on SPSS Statistics, let's take a few minutes to consider some of the central components of social science research. We will begin here by looking at the role of theory in conjunction with the social research process. In the following chapter, we will turn our attention to another fundamental aspect of scientific inquiry: measurement. The goal is not to make you an expert in the social research process but to give you the background necessary to master the techniques of data analysis presented in the remainder of this book.[4]

CONCEPTS AND THEORIES: DEPRIVATION THEORY

Given the variety of topics examined in social science research, no single, established set of procedures is always followed in social scientific inquiry. Nevertheless, data analysis almost always has a bigger purpose than the simple manipulation of numbers. Our larger aim is to learn something of general value about human social behavior. This commitment lies in the realm of theory. A primary goal of social scientific research is to examine the various concepts that constitute our knowledge of the social world and then develop them into theories that help us explain, understand, and make sense of the social world.

Concepts are general ideas or understandings that form the basis of social scientific research. Some of the social scientific concepts with which you are familiar might include social class, deviance, political orientations, prejudice, and alienation. The most useful concepts describe variations among people or groups. When thinking about social class, for example, we might distinguish the upper class, middle class, and working class, while the concept of prejudice leads us to consider those who are more prejudiced and those who are less prejudiced.

A ***theory*** is a statement or set of statements describing the relationships among concepts. Theories provide explanations about the patterns we find in human social life. Developing social theories is a

matter of discovering causal relationships between concepts, and we start to develop a theory by asking questions that help us begin to investigate those causal relationships. Examples of such questions include, "Does education reduce prejudice?" "Does gender affect how much people are paid?" "Are minority group members more liberal than majority group members?" and, "Are women more religious than men?" As for the last question, American social research has consistently shown that women are more religious than men, and the key concepts in that observed pattern are religiosity and gender. Because one of the subjects we are going to examine in this textbook is religiosity, we will begin with an example of a theory deriving from the sociology of religion.

The sociologists Glock, Ringer, and Babbie[5] developed what they called the deprivation theory of church involvement. Having asked why some church members participated more in their churches than did others, the researchers' analyses led them to conclude that those who were deprived of gratification (e.g., money, prestige, power, opportunities, and freedom) in secular society would be more likely to be active in church life than would those who enjoyed the rewards of secular society. In this case, the concepts under examination are deprivation, gratification, and church involvement. Some people are more deprived of gratification than others, and some people are more religiously involved than others. The research question is to find out if the degree to which people are deprived of secular gratification is somehow related to their degree of religious involvement.

Deprivation theory offers a plausible explanation of how the concepts of deprivation and religious involvement are related. It gives us a possible explanation—a theory—to help us make sense of why some people are more religious or more active in church than others. In this form, however, the concepts are too general for us to be able to test the theory empirically. Before we can test a theory, we must take the additional step of creating hypotheses. Unlike theories, well-developed hypotheses posit relationships between variables that are specific enough to permit testing.

In short, while theory is an important starting point in social science research, the empirical relationships predicted by the theory must be tested. To do that, we shift our focus from relationships between concepts to relationships between variables and, therefore, from theories to hypotheses.

VARIABLES AND HYPOTHESES: RELIGIOSITY

Variables are empirical indicators of the concepts we are researching. Variables, as their name implies, have the ability to take on two or more values. For instance, people can be classified in terms of their gender (male or female) or religious involvement (involved or not involved). When we identify empirical indicators for our concepts, they become variables.

A ***hypothesis*** is a statement of expectation derived from a theory that proposes a relationship between two or more variables. Specifically, a hypothesis is a tentative statement that proposes that variation in one variable "causes" or leads to variation in the other variable. We put *cause* in quotes here because more than a simple association is needed to attribute cause. To be a cause, a related variable must (among other requirements) precede the dependent variable in time and not be related to some other variable that is also related to the dependent variable. A ***dependent variable*** is the variable you are trying to explain (e.g., church involvement or religiosity), while an ***independent variable*** is the variable hypothesized to "cause," lead to, or explain variation in another variable (e.g., gender). We'll explore this further in Chapter 17.

Table 1.1 illustrates the differences between theories and hypotheses. Theories specify relationships between concepts in the world of ideas, while hypotheses specify expected relationships between variables in the world of empirical experiences.

Also, please note that relationships such as the one predicted in the hypotheses in Table 1.1 are ***probabilistic***, meaning that people displaying certain characteristics will have a higher probability of exhibiting certain behaviors but are not guaranteed to exhibit those behaviors. For example, Hypothesis B says that women, as a group, will have a higher average level of religious participation than will men, as a group. This does not mean that all women are more involved in religion than any men. It does mean, however, that if we asked men and women whether they attend church every week, a higher percentage of women than of men would say yes, even though some men would say yes, and some women would say no. That is the nature of probabilistic relationships.

TABLE 1.1 ⬢ Concepts, Theories, Variables, and Hypotheses

World of Ideas	Concepts
	1. Secular deprivation
	2. Religious involvement
	Theory: The more people experience secular deprivation, the more likely they will be religiously involved.
	Variables representing dimensions of secular deprivation
	a. Age
	b. Gender
	c. Socioeconomic status
World of Experiences	**Hypotheses**
	a. As people get older, their religious participation increases.
	Independent variable: Age
	Dependent variable: Religiosity
	b. Women will have greater religious participation than will men.
	Independent variable: Gender/sex
	Dependent variable: Religiosity
	c. The lower your income, the more likely you will be to participate in religious activities.
	Independent variable: Income
	Dependent variable: Religiosity

As Table 1.2 demonstrates, each variable contains two or more ***categories***, which are defined as the specific attributes that make up a variable. For instance, the categories of the variable *gender* are *male* and *female*, while the categories of the variable *social class* may be *upper class*, *middle class*, and *working class*.

TABLE 1.2 ⬢ Variables and Categories

Variables	Categories
Gender	Female
	Male
Religious involvement	Involved
	Not involved
Party identification	Democrat
	Independent
	Republican
Social class	Upper class
	Middle class
	Working class

The categories of each variable must meet two requirements: They should be both exhaustive and mutually exclusive. By ***exhaustive***, we mean comprehensive enough that it is possible to categorize every observation using those categories. Imagine, for instance, that you are conducting a survey, and one of your variables is religious affiliation. In order to measure respondents' religion, you devise a question that asks respondents simply, "What is your religion?" Let's say you give respondents only three choices: Protestant, Catholic, and Jewish. While most Americans would identify with one of these religious traditions, the categories certainly are not exhaustive. Muslims and Hindus, among others, would not find any of these categories to be descriptive of their traditions. To correct this problem, we would have to add more traditions, an "other" category, or both so that all respondents could fit themselves into at least one category. Moreover, we'd want a "none" category for those with no religious affiliation.

Mutually exclusive means every observation must fit into only one category. For instance, if we asked people for their religious affiliation and gave them the choices of Christian, Protestant, Catholic, and Jewish, then the categories would not be mutually exclusive since Protestants and Catholics would see themselves as being in the Christian category as well.

SOCIAL RESEARCH STRATEGIES: INDUCTIVE AND DEDUCTIVE

After developing a hypothesis, a researcher may decide to design and conduct a scientific study to test whether there is a relationship, such as the one proposed between gender and church involvement. Social scientists generally approach research in one of two ways: inductively or deductively.

In the previously mentioned study by Glock, Ringer, and Babbie, the researchers employed an ***inductive research*** strategy. First, they collected data regarding people's religious involvement and gender. After they completed their observations, they examined the data and constructed a theory to explain the relationships found among the variables.

An alternative and somewhat more common approach is ***deductive research***. Unlike inductive research, which begins with data collection, deductive research begins with social theory. A specific hypothesis is then deduced from the theory and tested to discover whether there is evidence to support it. To continue our example, the deprivation theory suggests that people who lack secular gratification will be more involved in religious activities. From that, we could derive the hypothesis that persons of lower socioeconomic status will attend church more often than will those of higher socioeconomic status. We could then collect data about people's socioeconomic status and church attendance and examine the data to see whether lower-status people really do attend church more than higher-status people. This would be considered deductive research because we began with the theory and tested a hypothesis with data.

Perhaps the simplest way to distinguish between inductive and deductive research approaches is by where they begin and end. While inductive research begins with data analysis and then moves to theory, deductive research begins with theory and then proceeds to data analysis and before returning to theory again. More simply, deduction can be seen as reasoning from general understandings to specific expectations, whereas induction can be seen as reasoning from specific observations to general explanations.

You can see, then, that whether an inductive or a deductive approach is used, the social research process involves many steps or stages.

Inductive Research

1. Collect data.
2. Analyze data.
3. *Induce* a theory to account for data.

Deductive Research

1. Find or formulate a theory.
2. *Deduce* hypotheses to test theory.
3. Collect data.
4. Analyze data.
5. Evaluate hypotheses.

While, in practice, the process of social research is not nearly as linear as these steps suggest, you can see that whether a researcher employs a deductive or an inductive strategy, the goal is always the same: to develop theories that help us explain, make sense of, and understand human social behavior.

The possible topics for exploration are, as you can imagine, endless. Whereas some social researchers are interested in understanding religiosity, others are interested in issues such as spousal abuse, child abuse, violence in schools, unemployment, political party identification, poverty, alcoholism, drug addiction, health care, crime, malnutrition, overpopulation, governmental corruption, and so on. The problems and issues of concern to social scientists are as manifold and complicated as human beings themselves. Despite the diversity of these questions and concerns, what connects social scientists is the belief that if used properly, the techniques and process of social science research can help us examine and begin to understand these complicated issues. Only when we understand what causes these problems, how they come about, and why they persist will we be able to solve them.

While the primary focus of this book is on one stage of the social research process—data analysis—we hope this book will inspire you to take some time to reflect on which of the many problems in contemporary life interest you. What issues or questions are you passionate about? What social problems or issues would you like to examine, understand, and potentially address?

THEORY AND RESEARCH IN PRACTICE

Now that we have focused a little on the relationship between theory and the social research process, let's examine some of the theoretical work that informs two of the many subjects we are going to analyze together in this book: political orientations and attitudes toward abortion.

Example 1: Political Orientations

One of the more familiar variables in social science is political orientation, which typically ranges from liberal to conservative. Political orientation lies at the heart of much voting behavior, and it also relates to a number of nonpolitical variables that you are going to discover for yourself shortly.

Each concept or variable we deal with in social research may have different aspects, which we call ***dimensions***. Three commonly examined dimensions of political orientation are (1) social attitudes, (2) economic attitudes, and (3) foreign policy attitudes. Let's examine each dimension briefly.

Some specific social attitudes and related behaviors might include abortion, same-sex marriage, and capital punishment. Let's see where liberals and conservatives generally stand on these issues:

Issue	Liberals	Conservatives
Abortion	Permissive	Restrictive
Same-sex marriage	Permissive	Restrictive
Capital punishment	Opposed	In favor of

In terms of economic issues, liberals are generally more supportive than conservatives of government programs such as unemployment insurance, welfare, and Medicare and of government economic regulation such as progressive taxation (the rich being taxed at higher rates), minimum-wage laws, and regulation of industry. By the same token, liberals are likely to be more supportive of labor unions than are conservatives.

Example 2: Attitudes Toward Abortion

Abortion is a social issue that has figured importantly in religious and political debates for years. The GSS contains several variables dealing with attitudes toward abortion. Each asks whether a woman should be allowed to get an abortion for a variety of reasons. The following list shows these reasons, along with the ***abbreviated variable names*** you'll be using for them in your analyses later on.

Abbreviated Variable Name	
ABDEFECT	Because there is a strong chance of a serious defect
ABNOMORE	Because a family wants no more children
ABHLTH	Because the woman's health would be seriously endangered
ABPOOR	Because a family is too poor to afford more children
ABRAPE	Because the pregnancy resulted from rape
ABSINGLE	Because the woman is unmarried
ABANY	Because the woman wants it, for any reason

Before we begin examining answers to the abortion attitude questions, it is worth taking a moment to reflect on their logical implications. Which of these items do you suppose would receive the least support? That is, which will have the smallest percentage of respondents agreeing with it? Think about that before continuing.

Logically, we should expect the smallest percentage to support ABANY because it contains all the others. For example, those who would support abortion in the case of rape might not support it for other reasons, such as the family's poverty. Those who support ABANY, however, would have to agree with both of those more specific items, plus all the rest.

Three of the items tap into reasons that would seem to excuse the pregnant woman from responsibility:

Abbreviated Variable Name	
ABDEFECT	Because there is a strong chance of a serious defect
ABHLTH	Because the woman's health would be seriously endangered
ABRAPE	Because the pregnancy resulted from rape

We might expect the highest percentages to agree with these items. We'll come back to this issue later to find out whether our expectations are correct.

When we analyze this topic using data, we will discover useful ways of measuring overall attitudes toward abortion. Once we've done that, we'll be in a position to find out why some people are generally supportive of abortion rights and why others are generally opposed.

CONCLUSION

This book has two educational aims. First, we want to share with you the excitement of social scientific research. You are going to learn that a table of numerical data, which may seem pretty boring on the

face of it, can hold within it the answers to many questions about why people think and act the way they do. Finding those answers requires that you learn some skills of logical inquiry. Second, we will show you how to use a computer program that is popular among social scientists: SPSS Statistics. It's the tool you will use to unlock the mysteries of society, just as a biologist might use a microscope or an astronomer might use a telescope.

Before getting started using SPSS Statistics, however, it is important that you have at least an initial appreciation for social research. In this chapter, we have focused in particular on the relationship between theory and the social research process. This examination will continue throughout the book. While most of our attention will focus on the skills involved in analyzing data, we will always want to make logical sense out of what we learn from our manipulations of the numbers. Measurement is a fundamental topic that bridges theory and research, so we turn our attention to that topic next.

Main Points

- The main purpose of this text is to introduce you to the logic and practice of social scientific research by showing you some simple tools you can use to analyze real-life data.
- Social and behavioral scientists' use of computer programs has evolved over many years because of the need to analyze large amounts of data.
- SPSS Statistics is a widely used, state-of-the-art statistical software program that will take you through all of the basics of using any sophisticated statistical package.
- A theory is a general statement or set of statements that describes and explains how different concepts are related to one another.
- A hypothesis is a tentative statement of expectation derived from a theory.
- A hypothesis proposes a relationship between two or more variables (independent and dependent variables) that can be tested by researchers employing scientific methods.
- The categories of variables must be both exhaustive and mutually exclusive.
- When a social scientist proceeds from theory to hypothesis development, data collection, and data analysis, the process is called deductive research.
- When a social scientist moves from data collection to data analysis and then induces a general theory based on those observations, the process is called inductive research.
- Theoretical work informs all of the subjects we are going to analyze in this book and, indeed, all of questions and issues of relevance to social scientists.

Key Terms

Review Questions

1. What are the two statistical packages most widely used by social scientists today?
2. Which of the versions of SPSS Statistics described is the least powerful in terms of the number of cases and variables it can handle? (Hint: It was discontinued by IBM after SPSS/PASW version 18, prior to its renaissance with version 22 and then discontinued once again after version 27.)
3. Which version of SPSS Statistics is the most powerful in this regard?
4. What version (or versions) of SPSS Statistics are you using?
5. Name two tasks for which one could use a statistical package, such as SPSS Statistics.
6. What is the primary goal of social scientific research?
7. Name two social scientific concepts.
8. What is the relationship between theories and hypotheses?
9. Complete the following statement: Theories are to concepts as hypotheses are to _____ _____.
10. Does a hypothesis propose a relationship between dimensions or variables?
11. The categories of each variable should meet what two requirements?
12. What, if anything, is the problem with the following categories of the variable *political views*: liberal and moderate? If there is a problem, how might you correct it?
13. What, if anything, is the problem with the following categories of the variable *political perspective*: liberal, Democrat, Republican, and conservative? If there is a problem, how might you correct it?
14. Construct a hypothesis based on the deprivation theory of church involvement using level of education as your independent variable.
15. List the categories of the variables you used to construct your hypothesis in response to Question 14.
16. Construct potential hypotheses to relate the following concepts, and identify the independent and dependent variable in each hypothesis. In addition, list the categories of each variable.

    Age and health

    Race and attitude toward affirmative action

    Gender and income
17. Which of the following is not a dependent variable: grade point average, church attendance, age, or number of children?
18. Which of the following is not a variable: occupation, amount of television viewing, female, or education level?
19. Consider the following hypothesis: People who earn more than $70,000 a year are more likely to vote Republican than people who earn less than $70,000 a year. Does this mean that all people who earn more than $70,000 a year vote Republican? Why, or why not?
20. Is the following statement true or false? A researcher who begins by collecting data and then develops a theory to explain their findings is engaged in deductive research.
21. A researcher formulates a hypothesis based on the "magic bullet theory" and then selects independent and dependent variables to test this hypothesis. What process is the researcher engaged in?
22. A researcher collects data on the spread of AIDS in the United States and then, based on their findings, develops a theory to explain why the rate of exposure and infection to the disease is higher among certain racial and ethnic groups than among others. In what process is the researcher engaged?

Notes

1. Earlier versions of SPSS Statistics for Windows or Macintosh may be used, but some of the instructions, procedures, and screens may be somewhat different from those in this book. For more details about specific computer and operating system requirements for various versions and configurations of SPSS, visit the IBM SPSS website: https://www.ibm.com/analytics/us/en/technology/spss.
2. While SPSS originally stood for and is still most commonly referred to as Statistical Package for the Social Sciences, SPSS Inc. recently said it had "updated

the meaning of the letters to more accurately reflect the company and its products. Today, SPSS stands for 'Statistical Product and Service Solutions.'" The PASW acronym stood for Predictive Analytics Software, and after SPSS Inc. became part of IBM in 2009, it was determined that the brand would become IBM SPSS Statistics, referred to as SPSS Statistics for short.

3. Throughout this book, we suggest various websites you may find useful. Keep in mind, however, that the World Wide Web is constantly changing. For this reason, some websites and online content referred to may no longer be available. If particular websites are no longer available, try using a search engine to find the information you need.
4. If you are thinking about designing a research study or just want to learn more about the process and practice of scientific inquiry, you may find the discussion in this book's last two chapters and accompanying appendices a useful starting point. You may also want to browse through the Reference section on the student study website at https://edge.sagepub.com/babbie11e for citations of texts that focus on the nature of social scientific inquiry, designing a research project, and other important aspects of the research process.
5. Glock, C. Y., Ringer, B. B., & Babbie, E. R. (1967). *To comfort and to challenge: A dilemma of the contemporary church*. Berkeley: University of California Press.

THE LOGIC OF MEASUREMENT

Measurement is one of the most fundamental elements of science. In the case of social research, the task of measurement is typically one of characterizing individuals in terms of the issues under study. Thus, a study of voting will characterize respondents in terms of the candidate for whom they plan to vote, and a study of abortion will describe people in terms of their attitudes on that topic.

VALIDITY PROBLEMS

Validity is a term used casually in everyday language, but it has a precise meaning in social research. It describes an indicator of a concept. Most simply, an indicator is said to be valid if it really measures the concept it is intended to measure; it is invalid if it doesn't.

As a simple example, let's consider political orientation, ranging from very liberal to very conservative. For an example of a clearly valid measure of this concept, here's the way the General Social Survey (GSS) asked about it.

> POLVIEWS: We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?
>
> 1. Extremely liberal
> 2. Liberal
> 3. Slightly liberal
> 4. Moderate, middle-of-the-road
> 5. Slightly conservative
> 6. Conservative
> 7. Extremely conservative

At the opposite extreme, a simple question about respondent gender would not be a valid measure of political orientation because political orientation and gender are different concepts. But now, let's consider another questionnaire item that does not come from the GSS. This item lies somewhere in between these two extremes of validity with regard to measuring political orientation.

> Question: Which of the two major political parties do you most identify with?
>
> 1. Democratic Party
> 2. Republican Party
> 3. Neither

This second item is another reasonable measure of political orientation. Moreover, it is related to the first because Democrats are, on the whole, more liberal than Republicans. On the other hand, there are conservative Democrats and liberal Republicans. If our purpose is to tap into the liberal–conservative dimension, the initial item that asks directly about political orientation is obviously a more valid indicator of the concept than is the item about political party.

This particular example offers us a clear choice as to the most valid indicator of the concept at hand, but matters are not always so clear-cut. If we were measuring levels of prejudice, for example, we could not simply ask, "How prejudiced are you?" both because no one is likely to admit to being prejudiced and because it's possible for someone to be prejudiced without knowing it. As we search for workable indicators of a concept such as prejudice, the matter of validity becomes something to which we must pay careful attention.

Validity is a concern not only when you collect and analyze your own data—a process known as ***primary research***—but also when you reanalyze data previously collected by someone else, as we do in this book. The process of reanalyzing someone else's data is referred to as ***secondary analysis***. Even if you can think of a survey question that would have captured your concept perfectly, the original researchers might not have asked it. Hence, you often need to use ingenuity in constructing measures that capture the quality in which you are interested. In the case of political orientation, for example, you might combine the responses to several questions—asking for attitudes about civil liberties, past voting behavior, political party identification, and so forth. We'll return to the use of multiple indicators shortly.

In large part, the question of validity is settled on *prima facie* grounds: We judge an indicator to be relatively valid or invalid on the face of it. It was on this basis that you had no trouble seeing that asking directly about political orientation was a valid indicator of that concept, whereas asking about a person's gender was definitely not a valid measure of political orientation. Later in this book, we'll explore some simple methodological techniques that are also used to test the validity of measures.

RELIABILITY PROBLEMS

Reliability is a different but equally important quality of measurement. ***Reliability*** refers to the quality of a measuring instrument that would cause it to report the same value in successive observations of a given case (provided the phenomenon being measured has not changed). For instance, if you step on a bathroom scale five times in a row and each time it gives you a different weight, the scale has a reliability problem. Conversely, if you step on a bathroom scale five times in a row and the scale gives you the same weight each time (even if the weight is wrong), the scale is reliable.

Similarly, if your statistics instructor administered the same test to you three times and each time you got a different score, even though your knowledge of statistics had not changed in the interim, the test would have a reliability problem. Conversely, if your instructor administered the same test three times and your score was the same each time, the test would be reliable.

In the context of survey research, reliability also concerns whether we can trust the answers that people give us even when their misstatements are honest ones. For instance, in medical research, some patients report in one survey that they have had a particular organ removed, only to indicate in subsequent surveys that they still have that organ.[1] Similarly, students of voting behavior regularly encounter individuals who claim on one survey that they did vote in the past presidential election and then claim in subsequent surveys that they either did not vote or do not remember whether they voted. As noted previously, these statements are often honest ones because it is difficult enough for most of us to recall what we did a few months ago, let alone several years ago.

Conceptually, the test of reliability is whether respondents would give the same answers repeatedly if the measurement could be made in such a way that (a) their situations had not changed (e.g., they hadn't had additional surgery to remove organs) and (b) they couldn't remember the answer they gave before.

As we suggested, empirically testing the reliability of an item requires multiple measures (e.g., your instructor must administer the statistics test at least two times in order to determine its reliability).

However, we can sometimes assess the reliability of a single item based on its practicality. For example, years ago, one of us was asked to help administer a survey to teenage drivers in California. Over researcher objections, the client insisted on asking the question, "How many miles have you driven?" and providing a space for the teenager to write in their response. Perhaps you can recognize the problem in this question by attempting to answer it yourself. Unless you have never driven an automobile, we doubt that you can report how many miles you have driven with any accuracy. In the survey mentioned, some teenagers reported driving hundreds of thousands of miles—highly questionable results for people who had only been driving for a few years. Unless they have been keeping a driving log their whole life, most people cannot report with any accuracy how many total miles they have driven.

In this situation, it would be better to provide respondents with a set of categories realistically reflecting the number of miles they are likely to have driven: fewer than 1,000 miles; 1,000 to 4,999 miles; 5,000 to 9,999 miles; and so on. Such a set of categories gives respondents a framework within which to place their own situations. Even though they still may not know exactly how much they have driven, there is a fair likelihood that the categories they choose will contain their correct answers. The success of this technique depends on our having a good idea in advance of what constitutes reasonable categories, perhaps as determined by previous research. As an alternative, we might ask respondents to volunteer the number of miles they have driven but limit the time period to something they are likely to remember. Thus, we might ask how many miles they drove during the preceding week or month.

DISTINGUISHING BETWEEN VALIDITY AND RELIABILITY

Perhaps the difference between validity and reliability can be seen most clearly in our previous example of weighing oneself on a bathroom scale. If you step on the scale repeatedly and it gives you a different weight each time, then the scale has a reliability problem. On the other hand, if the scale tells you that you weigh 125 pounds every time you step on it, then it's pretty reliable. However, if you actually weigh 225 pounds, then the scale, though reliable, has a problem in the validity department; it doesn't indicate your weight accurately.

Both validity and reliability are important in the analysis of data. If you are interested in learning why some people have deeply held religious beliefs and others do not, then asking people how often they attend church will be problematic. This question doesn't really provide a valid measure of the concept that interests you (religious belief) because anything you learn from asking the question will explain the causes of church attendance, not religious belief. Plus, if you asked people how many times they had attended church in the past year, any answers you received would probably not be reliable since, as with the driver survey mentioned previously, most people can't remember that far back. (It would be better, once again, to give them categories from which to choose.) Therefore, anything you might think you had learned about the causes of church attendance might be only a function of the errors people made in answering the question. You would have no assurance that another study would yield the same result, so your study would not be reliable.

MULTIPLE INDICATORS

Often, the solution to the problems just discussed lies in the creation of ***composite measures***, measurements of variables that combine two or more indicators of the variable, and the use of ***multiple indicators***, or several questions about the same concept. As a simple example, to measure the degree to which a sample group of church members holds Christian beliefs, you might ask the members of that group questions about several issues, each dealing with a particular belief:

- Belief in God
- Belief that Jesus was divine
- Belief in the existence of the Devil

- Belief in an afterlife: heaven and hell
- Belief in the literal truth of the Bible

The various answers respondents give to these questions could be used to create an overall measure of their Christian religious belief or lack thereof. In the simplest procedure, you could give respondents 1 point for each belief with which they agreed, allowing you to score them from 0 to 5 on the index. Notice that this is the same logic by which you may earn 1 point for each correct answer on an exam, with the total score being taken as an indication of how well you know the material.

Some social science concepts are implicitly multidimensional. Consider the concept of *social class*, for example. Typically, this term is used in reference to a combination of education, income, and occupation and, sometimes, dimensions such as social class identification and prestige. For the purpose of data analysis, this concept would therefore be measured through the use of multiple indicators. When it becomes appropriate in the analyses we are going to undertake together, we'll show you how to create and use some simple composite measures.

LEVELS OF MEASUREMENT

As we convert the concepts in our minds into empirical measurements in the form of variables, we sometimes have options as to those variables' level of statistical sophistication. Specifically, there are a number of different possibilities regarding the relationships among the categories constituting a variable. Therefore, in social research, we commonly speak of four ***levels of measurement***: nominal, ordinal, ratio, and interval.

Nominal Variables

Some variables simply distinguish different kinds of people. Gender is a good example of such a variable in the General Social Survey; the two categories of the variable, male and female, distinguish between men and women. Of course, such a treatment of this variable may not suit all research purposes, as transgender and intersex individuals are achieving more visibility, and there is more scholarly interest in them. The way we conceptualize and study gender will, no doubt, evolve in the coming years. Similarly, political party distinguishes Democrats from Republicans and from members of other parties; religious affiliation distinguishes Protestants from Catholics, Jews, and so forth. We refer to these variables as *nominal* because the categories constituting the variables are based on names.

Nominal variables simply name the different attributes constituting them. The attributes constituting a nominal variable (e.g., religious affiliation composed of Protestant, Catholic, Jewish, etc.) are simply different on their face. Republicans and Democrats are different from each other, as are Protestants and Catholics. In other cases, however, we can say more about the attributes making up variables.

Ordinal Variables

Many social scientific variables go a step beyond simply naming the different attributes constituting a variable. ***Ordinal variables*** arrange those attributes in some order: from low to high, from more to less, and so on. Whereas the nominal variable *religious affiliation* classifies people into different religious groups, *religiosity* might order them into groups such as very religious, somewhat religious, and not-at-all religious. And whereas the nominal variable *political party identification* simply distinguishes different groups (e.g., Democrats and Republicans), an ordinal measure of political philosophy might rank-order the very liberal, the somewhat liberal, the middle-of-the-road, the somewhat conservative, and the very conservative. Ordinal variables share the nominal variable's quality of distinguishing differences among people, but they add the quality of *rank-ordering* those differences.

At the same time, it is not meaningful to talk about the distances separating the attributes that make up an ordinal variable. For example, we have no basis for talking about the amount of liberalism separating the very liberal from the somewhat liberal or the somewhat liberal from the middle-of-the-road. We can say that the first group in each comparison is more liberal than the second, but we can't say by how much.

Ratio Variables

Some variables allow us to speak more precisely about the distances between the attributes constituting them. Consider age for a moment. The distance (in time) between 10 years old and 20 years old is exactly the same as that between 60 years old and 70 years old. Thus, it makes sense to talk about the distance (in time) between two ages (e.g., they are 10 years apart). Moreover, variables such as age have the quality of containing a genuine zero point—in this case, 0 years old. This quality is what allows us to examine ratios among the categories constituting such variables. Thus, we can say that a 20-year-old is twice as old as a 10-year-old. By comparison, we would have no grounds for saying one person is twice as religious as another, so religion can only be measured as a nominal or ordinal variable.

Ratio variables, then, share all of the qualities associated with nominal and ordinal variables, but they also have a zero point, and their categories can be separated by fixed units of distance and can thus be compared as ratios. Other examples of ratio measures include "annual amount of personal of household income," "years of schooling completed," and "hours worked per week."

Interval Variables

Somewhat rare in social research are ***interval variables***, which have the quality of standard intervals of measurement but lack a genuine zero point. One example is intelligence quotient, or IQ. Although IQ is calculated in such a way as to allow for a score of zero, such a score would not indicate a complete lack of any intelligence because the person would at least have been able to take the test.

Now, try to warm up to the idea of measuring temperature. The Celsius and Fahrenheit temperature scales both have 0° marks, but neither represents a total lack of heat, given that temperatures below zero are possible. The Kelvin scale, by contrast, is based on an absolute zero, which does represent a total lack of heat (measured in terms of molecular motion).

For most statistics used by social scientists, interval and ratio scales may be considered the same. When we start using SPSS Statistics, we'll see that its creators have lumped interval and ratio variables into a single category called ***scale***. Although these variables may be combined for practical purposes, the distinction between them helps us understand why a negative income might be interpreted as debt and why a negative age is impossible!

MEASUREMENT AND INFORMATION

Knowing a variable's level of measurement is important for selecting an appropriate statistic. Variables of different levels of measurement contain different amounts of information. The only information we have about nominal variables is the number of cases that share a common attribute. With ordinal variables, in addition to knowing how many cases fall into a given category, we know a greater-than, less-than relationship between the cases. Variables measured at the interval level have points that are equidistant from one another, so we know how much greater or less each case is than the others. Finally, with ratio variables, we have all of the characteristics of nominal, ordinal, and interval variables, plus the knowledge that zero is not arbitrary but means an absence of the phenomenon.

The statistics that SPSS Statistics has been programmed to compute are designed to make maximum use of the information preserved in a level of measurement. Using the mode on a sample of grade point averages ignores information used by the mean. Conversely, using the mean for a sample of religious preferences assumes information (equidistant points) not contained in a nominal measure. Responsible use of statistics requires selecting a statistic that matches the data's level of measurement. We'll talk about this more later. Right now, we want you to know that being able to identify a variable's level of measurement is essential for selecting the right statistical tool. We don't want to see you using a screwdriver when you need a hammer.

Table 2.1 displays the three primary levels of measurement that we discuss in this book: nominal, ordinal, and interval/ratio (scale). We purposefully designed the table as a series of steps to remind you that there is a hierarchy implied in the levels-of-measurement idea. Variables at the nominal level (the bottom step) contain the least amount of information, variables at the ordinal level (the middle step) contain more information, and variables at the interval/ratio level (the highest step) contain the most

TABLE 2.1 ⬢ Levels of Measurement

		INTERVAL/RATIO Distance between categories is meaningful. Examples: income (measured in thousands of dollars); age (measured in years)
	ORDINAL Categories can be rank ordered. Examples: social class (lower, working, middle, upper); attitudes toward gun control (strongly oppose, oppose, favor, strongly favor)	
NOMINAL Categories differ in name. Examples: gender (male, female); party identification (Democrat, Republican)		

information. You should also note that as you move from the nominal to the ordinal and finally to the interval/ratio level, each level has the qualities of the level(s) below it, plus a new trait.

MEASUREMENT OPTIONS

Sometimes, you will have options regarding the levels of measurement to be created in variables. For instance, although age can qualify as a ratio variable, it can be measured as ordinal (e.g., young, middle-aged, and old) or even as nominal (baby boomer or not baby boomer).

The significance of these levels of measurement will become more apparent when we begin to analyze the variables in our data set. As you'll discover, some analytic techniques are appropriate to nominal variables, some to ordinal variables, and some to ratio variables. On the one hand, you will need to know a variable's level of measurement to determine which analytic techniques are appropriate. On the other hand, where you have options for measurement, your choice of measurement level may be determined by the techniques you want to employ.

CLASSIFYING VARIABLES AS DISCRETE OR CONTINUOUS

In addition to distinguishing variables by their level of measurement, researchers sometimes classify variables as discrete or continuous. Just as distinguishing between levels of measurement helps us choose appropriate statistics, so too does knowing whether variables are discrete or continuous.

Discrete variables are variables whose values are completely separate from one another. In the GSS data, these include variables such as NUMBER OF SIBLINGS and SEX. These are variables with a limited number of distinct (i.e., discrete) values or categories that cannot be reduced or subdivided into smaller units or numbers. Discrete variables can be nominal (sex), ordinal (class rank), or interval/ratio (number of siblings). All of these variables are discrete because their values cannot be subdivided or reduced. A respondent may, for instance, have 1 dog, but they cannot have 0.5 or 0.25 dogs. Dogs, like people, come in discrete units of 1; they simply cannot be subdivided into smaller units.

(Note that it is possible to have an average number of dogs per family in a town that is not a whole number, say 1.5 dogs. This is an average; it does not and cannot indicate that any single family has

a fractional physical dog—dogs, like people, are whole creatures only, like Mina in the photo to the right posing with a previous edition of *Adventures in Social Research*!)

Continuous variables, on the other hand, are variables whose values can be infinitely subdivided, such as AGE or EDUC (education measured in years). These variables are continuous both because they have a time dimension and because time can be infinitely subdivided (i.e., years, months, weeks, days, hours, minutes, seconds, etc.). The level of measurement for continuous variables can be either interval/ratio (age measured in years) or ordinal (age measured as infant, toddler, adolescent, preteen, teenager, etc.).

Unlike levels of measurement, which tell us the amount of information provided by a measure, the discrete versus continuous distinction gives us information about the underlying characteristics of a variable. In particular, it refers to the phenomenon's divisibility, or whether the values of a variable can be subdivided into ever-smaller units, as described in Table 2.2.

Source: Photo by Billy Wagner.

TABLE 2.2 ⬢ Tips for Distinguishing Between Discrete and Continuous Variables

Discrete Variable	Example: Number of Siblings
	• Units CANNOT be reduced to ever-smaller units.
	• There is NOT an infinite number of other possible categories between any two categories of this variable (i.e., 1 and 2 siblings, or 2 and 3 siblings, etc.).
	• Variables can be nominal, ordinal, or interval/ratio (scale).
Continuous Variable	**Example: Age Measured in Years**
	• Units CAN be further reduced to smaller units (i.e., months, weeks, days, hours, minutes, seconds, etc.).
	• There ARE an infinite number of other possible categories between any two categories of this variable (i.e., 19 and 20 years old, 20 and 21 years old, etc.).
	• Variables can be interval or ratio (scale).

CONCLUSION

Measurement is a fundamental aspect of social science research. It may be seen as the transition from concepts to variables—from sometimes ambiguous mental images to precise, empirical measures. Whereas we often speak casually about such concepts as prejudice, social class, and liberalism in everyday conversation, social scientists must be more precise in their use of these terms.

Chapters 1 and 2 have given you a brief overview of two important issues in social scientific inquiry that are directly relevant to our primary focus—computerized statistical analysis. The chapters that follow build on this discussion of theory and measurement and show you the concrete techniques you need to engage in data analysis.

Main Points

- Measurement is a vital component of social scientific research.
- In designing and evaluating measurements, social scientists must pay particular attention to the problems of validity and reliability.
- A common remedy for problems of validity and reliability is the construction of composite measures using multiple indicators.
- Level of measurement signifies the different amounts and types of information obtained about a variable, and consideration of level of measurement is essential for selecting appropriate statistical tools.
- The four levels of measurement are nominal, ordinal, ratio, and interval.
- Variables of different levels of measurement contain different amounts of information.
- There is an implied hierarchy of the various levels of measurement, with the nominal level being at the bottom (meaning it can provide the least information), the ordinal level in the middle, and the interval and ratio levels being at the top (meaning they can provide the most information).
- In addition to classifying variables by their level of measurement, researchers distinguish between discrete and continuous variables.
- Discrete variables cannot be infinitely subdivided into ever-smaller units, whereas continuous variables can.

Key Terms

Composite measures 17
Continuous variables 21
Discrete variables 20
Interval variables 19
Levels of measurement 18
Multiple indicators 17
Nominal variables 18
Ordinal variables 18
Primary research 16
Ratio variables 19
Reliability 16
Scale 19
Secondary analysis 16
Validity 15

Review Questions

1. A researcher sets out to measure drug use on U.S. college campuses by asking a representative sample of undergraduates whether they are currently receiving federal grants or loans. What is the problem with this measure?
2. A researcher asks a representative sample of baby boomers how much alcohol they consumed during their college years and includes on the survey form a space for them to write in the actual number of drinks they consumed. One month later, the researcher administers the same questionnaire to the same respondents, and more than half the respondents report consuming much less alcohol during their college years than they reported just a month earlier. What is the problem with this measure?
3. Multiple indicators are useful in dealing with what types of problems?
4. Name one reason why it is important to know or be able to identify a variable's level of measurement.
5. Ordinal variables have all the qualities of variables at which other level of measurement?
6. Ratio variables have all the qualities of variables at which other levels of measurement?
7. The creators of SPSS Statistics have combined ratio and interval variables into one category that they refer to as _____.
8. A variable whose values can be infinitely subdivided is called a _____ variable.

Identify the level of measurement of each of the following variables, and classify the variables as either discrete or continuous (Questions 9–11).

9. A researcher measures respondents' attitudes toward premarital sex by asking the following question: "If a man and woman have sexual relations before marriage, do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?"
10. A researcher measures the amount of television viewing by asking the following question: "On the average day, how many hours do you personally watch television?" The researcher then asks respondents to fill in the actual number of hours in a space provided.
11. A researcher measures marital status by asking respondents whether they are currently married, widowed, divorced, separated, or never married.

Indicate whether the following statements (Questions 12–17) are true or false.

12. Certain variables can be measured at both the nominal and ordinal levels.
13. You are invited to watch a screening of a new movie and then rate it as excellent, good, fair, or poor. The level of measurement is nominal because the ratings differ in name.
14. A researcher asks respondents to indicate the last four digits of their Social Security numbers. The level of measurement for this variable is interval/ratio because the distance between categories is meaningful.
15. A researcher asks respondents how many siblings they have. This variable can be categorized as continuous.
16. A researcher asks respondents how long they have lived at their current residence. This variable can be categorized as continuous.
17. Discrete variables can be either nominal or ordinal but not interval/ratio.

Complete the activities below (Questions 18–20).

18. Construct measures of *annual income* at two levels of measurement.
19. Classify the variables you constructed in response to Question 18 as either discrete or continuous.
20. Construct measures of *individual age* at two levels of measurement.

Note

1. The authors are grateful to Professor Randall MacIntosh, California State University–Sacramento, for this suggestion.

DESCRIPTION OF DATA SETS

The General Social Survey

The data we provide for your use here are real. They come from the responses of 2,348 adult Americans selected as a representative sample of the United States in 2018. These data are a major resource for professional social scientists and are the basis of many published books and articles.

The ***General Social Survey (GSS)*** is conducted regularly by the ***National Opinion Research Center (NORC)*** in Chicago, with financial support from the National Science Foundation and private sources. The purpose of the GSS program is to provide the nation's social scientists with accurate data for analysis.

This survey was the brainchild of Jim Davis (1929–2016). The GSS, which began in 1972, was conducted annually until 1994, when it became biennial.[1] Over the past 50 years, the GSS has asked more than 50,000 respondents more than 5,000 questions on topics ranging from attitudes toward abortion to star signs in the zodiac. In recent years, the GSS asked questions on issues such as Internet use, religious transformations, spirituality, genetic testing, heredity, stress, violence in the workplace, immigration, altruism, alcohol consumption, sexual behavior, social networks, and group membership.

While this chapter provides you with a brief overview of some of the central components of the GSS, you can access further information about the GSS and other data by visiting one of the following websites:

- NORC is a nonprofit corporation, affiliated with the University of Chicago, that conducts the GSS. Visit the NORC home page at http://www.norc.org. Once there, click on *Projects* and filter on *General Social Survey* or go directly to http://gss.norc.org. The GSS Data Explorer website is particularly useful for examining and obtaining GSS data: https://gssdataexplorer.norc.org.
- The Inter-University Consortium for Political and Social Research (ICPSR) is associated with the University of Michigan. Part of its mission is to maintain and provide access to a vast archive of social science data for research and instruction. ICPSR's home page can be found at http://www.icpsr.umich.edu.
- The Roper Center for Public Opinion Research is an archive of public opinion research, affiliated with the Cornell University (formerly affiliated with the University of Connecticut), that provides access to the GSS. The Roper Center home page is at https://ropercenter.cornell.edu. Click on *Search Datasets* and enter "General Social Survey," or visit https://ropercenter.cornell.edu/general-social-survey.
- You can also access this information online by using your favorite search engine. Simply search using one of the following terms: *General Social Survey (GSS), GSS 2016, National Opinion Research Center (NORC), Inter-University Consortium for Political and Social Research (ICPSR*), or *Roper Center for Public Opinion Research*. If you do such a search, you will find that Canada also has a GSS sponsored by Statistics Canada, the equivalent of the United States Census Bureau.

SAMPLING

The data provided by the GSS are a representative sample of American adults. This means that anything we learn about the 2,348 people sampled can be taken as an accurate reflection of what all (noninstitutionalized, English-speaking) American adults (18 years of age or older) would have said if we could have interviewed them all.

Creating a truly representative sample requires developing a selection technique that ensures every person in the population has an equal probability of being included. Since the previous GSS, the U.S. Postal Service and the U.S. Census Bureau have made available lists of household units representing 72% of the U.S. population. With the aid of computers, random samples of households are drawn from the lists for inclusion in the GSS sample. For the 28% of the population not on household lists, Census tracts and enumeration districts are selected, and then, housing units are randomly drawn for inclusion.

This complex and sophisticated sampling process makes it possible for the responses of 2,348 individuals to provide an accurate reflection of the feelings of all adult Americans. The U.S. Census Bureau uses similar techniques for the purpose of government planning, as do polling firms that predict voting behavior with relative accuracy.

For a detailed description of the sample design the GSS uses, see "Appendix A: Sampling Design and Weighting" in the *General Social Surveys 1972–2018 Cumulative Codebook*.[2]

The *Adventures in Social Research* student study site (https://edge.sagepub.com/babbie11e) includes three data files. One, named ***GSS2018.SAV***, includes all 2,348 cases across 1,065 variables from the GSS. We have also created a more compact data file, ***ADVENTURES.SAV***, which includes all 2,348 respondents but only 143 variables, the ones we use in the book along with other relevant variables that lend themselves to analysis that might be appropriate in the open-ended exercises in this book. We also created another file, ***STUDENTS.SAV***, by reducing the size of the GSS sample to 1,500 cases (through random subselection) so that those who might be using the student version of SPSS (for both Windows and Macintosh) can access some of the GSS 2018 data. We were limited to 50 variables, so we had to pick and choose. As we noted in Chapter 1, whereas the professional and graduate versions of SPSS Statistics for Windows and Macintosh are, for all practical purposes, limited only by the size of the computer on which they are installed, the student version for both operating systems is limited to 1,500 cases and 50 variables.

We have selected 143 of the variables from the GSS to include in the ADVENTURES.SAV data file. The ADVENTURES.SAV data file contains the data that are used in the examples in each of the chapters, demonstrating the techniques and analysis. In addition, the ADVENTURES.SAV file contains the data needed to answer the Laboratory Assignment questions at the end of each chapter, starting with Chapter 4.

The data you have at hand, then, can be taken as an accurate reflection of the characteristics, attitudes, and behaviors of noninstitutionalized Americans 18 years of age and older in 2018. This statement does need to be qualified slightly, however. When you analyze the data and learn that 50.1% (variable: ABANY, GSS, 2018) of the sample said that they supported a woman's unrestricted right to have an abortion for any reason, you are safe in assuming that just over 50% of the entire U.S. adult population feels that way. Because the data are based on a sample rather than on asking everyone, however, we need to anticipate some degree of sampling error. You can think of ***sampling error*** as the extent to which the responses of those sampled differ from the responses of the larger population (English-speaking persons 18 years of age and older living in noninstitutionalized arrangements within the United States in 2018). As a general rule, the greater the sampling error, the less representative the sample. It would not be strange, based on the example given, to discover that 48% to 52% of the total adult population (rather than precisely 50%) support a woman's unrestricted right to have an abortion for any reason. It is inconceivable, however, that as few as 10% or as many as 90% hold the opinion in question.

As a rough guideline, you can assume that the sampling error in this data set is plus or minus only a few percentage points. Later in the book, we'll talk about how to calculate sampling error for specific pieces of data.

Even granting the possibility of sampling error, however, our best estimate of what's true among the total U.S. population is what we have learned from probability samples. Thus, if you had to bet on the percentage of the total U.S. population who support a woman's unrestricted right to an abortion, you should put your money on 50%. You would be better off, however, betting on the range of 48% to 52%.

DATA COLLECTION

The GSS data were collected in face-to-face household interviews. Once the sample households were selected, professional interviewers were dispatched to call on each one. The interviewers asked each of the questions and wrote down the respondents' answers. Each interview took about 90 minutes.

To maximize the amount of information that can be collected in this massive interviewing project, the GSS uses a split-ballot design, such that NORC asked some questions of a random subsample of the households and other questions of the other households. Still other questions were asked of all respondents. When we begin analyzing the GSS data, you will notice that some data items have a substantial number of respondents marked ***missing data***. For the most part, this term refers to respondents who were not asked that particular question.

Although some of the questions were posed to only a subsample of households, you can still take the responses as representative of the U.S. adult population, except that the degree of sampling error, mentioned previously, is larger.

For more information about how the GSS data were collected, see "Appendix B: Field Work and Interviewer Specifications" and "Appendix C: General Coding Instructions" in the *General Social Surveys 1972–2018 Cumulative Codebook*.

THE CODEBOOK: APPENDIX A

The questionnaire items included in the files you will be using throughout this text (ADVENTURES.SAV) are listed in Appendix A of this book. We attempted to choose variables that are not only interesting but also relevant to students from a variety of social and health science disciplines, including communications, criminal justice, education, health policy, political science, public administration, public health, social work, and sociology.

Before proceeding to the next chapter, you may want to take a few minutes to review the variables in Appendix A. Before long, you'll be getting much more familiar with them. As you analyze survey data, it is important to know exactly how questions were asked if you are to understand the meaning of the answers given in response. Appendix A includes the following information for each of the variables contained in your ADVENTURES.SAV file.

- Abbreviated variable names—used by SPSS Statistics to access variables
- Question wording—how the interviewer asked the question
- ***Values***—sometimes called *numeric values* or *numeric codes*, used to code responses
- ***Value labels***—used to identify categories represented by values. Please note that in Appendix A, we have excluded the following value labels: ***NA (no answer)***, ***DK (don't know)***, and ***IAP (inapplicable)***. These labels refer to cases when the respondent offered "no answer," the respondent said they "did not know," or the question was not asked and, thus, is "inapplicable."
- Level of measurement for each variable

Data Set 1: ADVENTURES.SAV

ADVENTURES.SAV will be used and referred to in the demonstrations in the body of the chapters that follow this one. You can follow along with these examples on your own computer. This file contains a all 2,348 respondent cases and 143 variables drawn from the 2018 GSS.

Data Set 2: STUDENTS.SAV

STUDENTS.SAV is made available for students who may only have access to the Student Version of SPSS, no longer supported by SPSS. (Note that the Grad Pack Version of SPSS can support the ADVENTURES.SAV file). This STUDENTS.SAV file contains a randomly selected subsample of 1,500 cases and 50 variables drawn from the 2018 GSS.

Data Set 3: GSS2018.SAV

GSS2018.SAV is the full 2018 GSS file. It includes all 2,348 cases, as well as all 1,065 original variables. The *GSS Cumulative Codebook (1972-2018)*, also provided on the *Adventures in Social Research, 11th edition* website, provides details for each of the variables in the GSS2018.SAV file.

CONCLUSION

After reading this chapter and Appendix A, you should be familiar with the General Social Survey and the variables with which you will be working. The data you will be using are real and can be taken as an accurate reflection of the attitudes, opinions, beliefs, characteristics, and behaviors of adult Americans in 2018.

In the next chapter, we are going to get started using SPSS Statistics. Once you have learned how to launch an SPSS Statistics session and access the files on the *Adventures in Social Research* website (https://edge.sagepub.com/babbie11e), you will be ready to begin exploring your data. With the help of SPSS Statistics and some simple tools, we think you will find that the possibilities for discovery are both rich and rewarding.

Main Points

- The General Social Survey (GSS) is a national survey of adult Americans that has been conducted annually or every other year since 1972.
- The 2018 data are based on a sample of 2,348 adult Americans and can be taken as an accurate reflection of noninstitutionalized, English-speaking adult Americans in 2018.
- The data were collected in face-to-face household interviews averaging about 90 minutes each.
- The website associated with this book (https://edge.sagepub.com/babbie11e) contains three data files: ADVENTURES.SAV, STUDENTS.SAV, and GSS2018.SAV.
- ADVENTURES.SAV is the file you will be using as you work your way through the chapters and to complete the exercises at the end of each chapter.
- The ADVENTURES.SAV data file contains data from all 2,348 respondents across 143 variables, taken from the full GSS 2018 data set.

Key Terms