

LEARNING MADE EASY



4th Edition

SPSS[®] Statistics

for
dummies[®]
A Wiley Brand

Perform all new
statistical functions

Write, edit and format
SPSS Syntax faster

Understand user interface
enhancements

Jesus Salcedo
Keith McCormick



SPSS[®] Statistics

4th Edition

**by Jesus Salcedo
and Keith McCormick**

**for
dummies[®]**
A Wiley Brand

SPSS® Statistics For Dummies®, 4th Edition

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2020 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and may not be used without written permission. SPSS is a registered trademark of International Business Machines Corporation. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit <https://hub.wiley.com/community/support/dummies>.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2020941201

ISBN 978-1-119-56083-8 (pbk); ISBN 978-1-119-56087-6 (ebk); ISBN 978-1-119-56082-1 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

Introduction	1
Part 1: Getting Started with SPSS	5
CHAPTER 1: Introducing SPSS	7
CHAPTER 2: Finding the Best SPSS for You	17
CHAPTER 3: Getting to Know SPSS by Running a Simple Session	25
Part 2: Getting Data into and out of SPSS	43
CHAPTER 4: Understanding SPSS Data: Defining Metadata	45
CHAPTER 5: Opening Data Files	59
CHAPTER 6: Getting Data and Results from SPSS	71
CHAPTER 7: More about Defining Your Data	81
Part 3: Messing with Data in SPSS	95
CHAPTER 8: The Transform and Data Menus	97
CHAPTER 9: Computing New Variables	119
CHAPTER 10: Some Useful Functions	133
CHAPTER 11: Combining Files	147
Part 4: Graphing Data	161
CHAPTER 12: On the Menu: Graphing Choices in SPSS	163
CHAPTER 13: Building Graphs Using Chart Builder	183
Part 5: Analyzing Data	211
CHAPTER 14: Using Descriptive Statistics	213
CHAPTER 15: Knowing When Not to Trust Your Data	227
CHAPTER 16: Testing One Group	239
CHAPTER 17: Showing Relationships between Categorical Variables	251
CHAPTER 18: Showing Relationships between Continuous Dependent and Categorical Independent Variables	267
CHAPTER 19: Showing Relationships between Continuous Variables	285
Part 6: Getting More Advanced with Analyzing Data	301
CHAPTER 20: Doing More Advanced Analyses	303
CHAPTER 21: What Is Normal Anyway?	327
CHAPTER 22: When to Do What	345

Part 7: Making SPSS Your Own	351
CHAPTER 23: Changing Settings	353
CHAPTER 24: Editing Charts and Chart Templates	371
CHAPTER 25: Editing Tables	383
Part 8: Programming SPSS with Command Syntax	393
CHAPTER 26: Getting Acquainted with Syntax	395
CHAPTER 27: Adding Syntax to Your Toolkit	403
Part 9: The Part of Tens	413
CHAPTER 28: Ten (or So) Modules You Can Add to SPSS	415
CHAPTER 29: Ten Useful SPSS Resources	425
CHAPTER 30: Ten SPSS Statistics Gotchas	433
Index	443

Table of Contents

INTRODUCTION	1
About This Book	1
About the Fourth Edition	2
Foolish Assumptions	2
Icons Used in This Book	3
Beyond the Book	3
Where to Go from Here	3
 PART 1: GETTING STARTED WITH SPSS	 5
CHAPTER 1: Introducing SPSS	7
SPSS's Job, Our Job, and Your Job	7
SPSS's job	8
Our job	8
Your job	9
Garbage In, Garbage Out: Recognizing the Importance of Good Data	9
Talking to SPSS: Can You Hear Me Now?	12
The graphical user interface	12
Syntax	12
Programmability	13
How SPSS works	13
Getting Help When You Need It	15
 CHAPTER 2: Finding the Best SPSS for You	 17
Campus Editions	19
Subscription Plans	20
Commercial Editions	22
What's New in Version 27	24
 CHAPTER 3: Getting to Know SPSS by Running a Simple Session	 25
Opening a Dataset	25
Running an Analysis	27
Interpreting Results	30
Creating Graphs	33
Investigating Data	37

PART 2: GETTING DATA INTO AND OUT OF SPSS	43
CHAPTER 4: Understanding SPSS Data: Defining Metadata	45
Entering Variable Definitions on the Variable View Tab	46
Name	47
Type	47
Width	51
Decimals	52
Label	52
Values	53
Missing	54
Columns	55
Align	55
Measure	55
Role	56
Entering and Viewing Data Items on the Data View Tab	56
CHAPTER 5: Opening Data Files	59
Getting Acquainted with the SPSS File Format	59
Reading Simple Data from a Text File	60
Transferring Data from Another Program	65
Reading an Excel file	67
Reading from an unknown program type	68
Saving Data	69
CHAPTER 6: Getting Data and Results from SPSS	71
Exporting Data to Another Program	71
Navigating SPSS Statistics Viewer	72
Moving SPSS Output to Other Applications	74
Copying and pasting output	74
Exporting output	75
Printing Data	80
CHAPTER 7: More about Defining Your Data	81
Working with Dates and Times	82
Using the Date and Time Wizard	84
Creating and Using a Multiple-Response Set	86
Copying Data Properties	90
PART 3: MESSING WITH DATA IN SPSS	95
CHAPTER 8: The Transform and Data Menus	97
Sorting Cases	97
Selecting the Data You Want to Look At	100
Splitting Data for Easier Analysis	103

Counting Case Occurrences.....	104
Recoding Variables	107
Recoding into different variables	107
Automatic recoding.....	110
Binning	113
Optimal Binning.....	117
CHAPTER 9: Computing New Variables	119
Calculating a New Variable with a Formula.....	120
Calculating a New Variable with a Condition.....	122
Using System Variables.....	124
Contrasting \$Sysmis with SYSMIS.....	125
Understanding Missing Data in Formulas	127
Efficiently Calculating with Multiple Formulas	129
CHAPTER 10: Some Useful Functions	133
The LENGTH Function.....	134
The ANY Function	137
The MEAN Function and Missing Data	139
RND, TRUNC, and MOD	141
Logicals, the MISSING Function, and the NOT Function	143
String Parsing and Nesting Functions.....	144
Calculating Lags.....	146
CHAPTER 11: Combining Files	147
Merging Files by Adding Cases	147
Merging Files by Adding Variables	152
PART 4: GRAPHING DATA.....	161
CHAPTER 12: On the Menu: Graphing Choices in SPSS.....	163
Building Graphs the Chart Builder Way	164
The Gallery tab.....	164
The Basic Elements tab.....	168
The Groups/Point ID tab.....	169
The Titles/Footnotes tab.....	170
The Element Properties tab.....	170
The Chart Appearance tab.....	176
The Options tab.....	177
Building Graphs with Graphboard Template Chooser	178
CHAPTER 13: Building Graphs Using Chart Builder.....	183
Simple Graphs	184
Simple scatterplots	184
Simple dot plots.....	185

Simple bar graphs	186
Simple error bars.	187
Simple histograms.	189
Population pyramids.	191
Stacked area charts.	192
Fancy Graphs	194
Charts with multiple lines.	194
Colored scatterplots	196
Scatterplot matrices	198
Stacked bar charts.	199
Pie charts	200
Clustered range bar graphs.	202
Differenced area graphs.	202
Dual-axis graph	204
Fancy Maps Using Graphboard Template Chooser	205
Heat map	206
Choropleth of values.	206
Coordinates on a reference map	209
PART 5: ANALYZING DATA.	211
CHAPTER 14: Using Descriptive Statistics	213
Looking at Levels of Measurement.	213
Defining the four levels of measurement	214
Defining summary statistics.	215
Focusing on Frequencies for Categorical Variables	217
Understanding Frequencies for Continuous Variables	221
Summarizing Continuous Variables with the Descriptives Procedure.	224
CHAPTER 15: Knowing When Not to Trust Your Data	227
Sampling.	227
Understanding Sample Size	228
Testing Hypotheses.	229
Calculating Confidence Intervals.	231
Conducting In-Depth Hypothesis Testing	232
Using the Normal Distribution	235
Working with Z-Scores	236
CHAPTER 16: Testing One Group	239
Conducting Inferential Tests	239
Running the Chi-Square Goodness of Fit Test.	240
Running the One-Sample T-Test Procedure	246

CHAPTER 17: Showing Relationships between Categorical Variables	251
Running the Crosstabs Procedure	252
Running the Chi-Square Test of Independence	256
Comparing Column Proportions	260
Adding Control Variables	261
Creating a Clustered Bar Chart	264
CHAPTER 18: Showing Relationships between Continuous Dependent and Categorical Independent Variables	267
Conducting Inferential Tests	268
Using the Compare Means Dialog	268
Running the Independent-Samples T-Test Procedure	269
Comparing the Means Graphically	275
Running the Summary Independent-Samples T-Test Procedure	277
Running the Paired-Samples T-Test Procedure	280
CHAPTER 19: Showing Relationships between Continuous Variables	285
Viewing Relationships	286
Running the Bivariate Procedure	288
Running the Simple Linear Regression Procedure	292
PART 6: GETTING MORE ADVANCED WITH ANALYZING DATA	301
CHAPTER 20: Doing More Advanced Analyses	303
Running the One-Way ANOVA Procedure	303
Conducting Post Hoc Tests	311
Comparing Means Graphically	314
Running the Multiple Linear Regression Procedure	315
Viewing Relationships	325
CHAPTER 21: What Is Normal Anyway?	327
Understanding Nonparametric Tests	328
Understanding Distributions	328
Running a Nonparametric Independent Samples Test	331
Running a Nonparametric Related Samples Test	338
CHAPTER 22: When to Do What	345
Determining Which Statistical Test to Perform	346
Using Advanced Techniques	350

PART 7: MAKING SPSS YOUR OWN	351
CHAPTER 23: Changing Settings	353
General Options	354
Language Options	356
Viewer Options	357
Data Options	358
Currency Options	360
Output Options	361
Chart Options	362
Pivot Tables Options	364
File Locations Options	365
Scripts Options	366
Multiple Imputations Options	368
Syntax Editor Options	369
Privacy Options	370
CHAPTER 24: Editing Charts and Chart Templates	371
Changing and Editing Axes	372
Changing the axis range	372
Scaling the axis range	373
Changing Style: Lines and Symbols	376
Editing chart lines	376
Editing data points	378
Applying Templates	380
CHAPTER 25: Editing Tables	383
Working with TableLooks	384
Style Output	387
Pivoting Trays	390
PART 8: PROGRAMMING SPSS WITH COMMAND SYNTAX	393
CHAPTER 26: Getting Acquainted with Syntax	395
Pasting	396
Performing a Series of Related Compute Statements	399
Labeling	400
Repeatedly Generating the Same Report	400
CHAPTER 27: Adding Syntax to Your Toolkit	403
Your Wish Is My Command	404
Understanding Keywords	405
Declaring Data	406
Commenting Your Way to Clarity	407

Running Your Code	408
Controlling Flow and Executing Conditionals	410
IF	410
DO IF	411
SELECT IF	412
PART 9: THE PART OF TENS	413
CHAPTER 28: Ten (or So) Modules You Can Add to SPSS	415
The Advanced Statistics Module	416
The Custom Tables Module	416
The Regression Module	418
The Categories Module	418
The Data Preparation Module	419
The Decision Trees Module	419
The Forecasting Module	420
The Missing Values Module	421
The Bootstrapping Module	421
The Complex Samples Module	422
The Conjoint Module	422
The Direct Marketing Module	422
The Exact Tests Module	423
The Neural Networks Module	424
CHAPTER 29: Ten Useful SPSS Resources	425
Supporting Websites for This Book	425
LinkedIn and LinkedIn Groups	426
IBM SPSS Statistics Certification	427
IBM Data Science Community	427
SPSSX-L	427
Online Videos	428
Twitter	429
Live Instruction	430
Asynchronous Instruction and Tutorials	431
SPSS Statistics for Data Analysis and Visualization	432
CHAPTER 30: Ten SPSS Statistics Gotchas	433
Failing to Declare Level of Measurement	433
Conflating String Values with Labels	434
Failing to Declare Missing Data	435
Failing to Find Add-on Modules and Plug-Ins	435
Failing to Meet Statistical and Software Assumptions	437

Confusing Pasting Syntax with Copy and Paste	438
Thinking You Create Variables in SPSS as You Do in Excel	438
Getting Confused by Listwise Deletion.	439
Losing Track of Your Active Dataset	440
Forgetting to Turn Off Select and Split and Weight	441
INDEX	443

Introduction

Good news! You don't have to know diddlysquat about the math behind statistics to be able to come up with well-calculated conclusions and display them in fancy graphs. You won't be doing any calculations by hand. All you need is the IBM SPSS Statistics software and a bunch of numbers. This book shows you how to type the numbers, select options in the menus, and produce brilliant statistics. And interpret them properly, too! It really is as simple as that.

About This Book

Some sections of the book are written as stand-alone tutorials to make it easy for you to get into whatever you're after. After you're up and running with SPSS, you can skip around and read just the sections you need. You really don't need to read straight through the entire book. However, the book's chapters are organized in parts, and it's generally a good idea to read the chapters in a part in sequence. And you should start with Part 1 if you are brand new to SPSS.

This book is not about math. It's about statistics. You don't derive anything. You don't do any math by hand or look up numbers in statistical tables. You won't find one explanation of how calculations are performed under the hood. This book is about the things you can do to command SPSS to calculate statistics for you. You can avoid all of those fancy formulas with lots of sigmas in them and still use SPSS to produce some nifty stats!

However, if you decide to study the techniques of statistical calculation, you'll be able to understand what SPSS does to produce numbers. Your main advantage in understanding the process to that degree of detail is that you'll be able to choose a calculation method that more closely models the reality you're trying to analyze. Although SPSS rescues you from the stress that some associate with manual calculations, it still leaves it up to you to interpret the results. We spend a lot of pages on carefully preparing you for that responsibility.

Throughout the book are examples that use data stored in files. These files are freely available to you. Some files are installed with IBM SPSS Statistics in the SPSS installation directory, which is `\Program Files\SPSS` (unless you chose

another location during installation). Other files are available via the Welcome dialog when you launch SPSS. Most of the files, however, were designed for this book and are available on the book's companion website (see "Beyond the Book" for more information). In every case, the files were carefully curated to demonstrate some specific capability of SPSS.

And a technical note: The official name of the product is IBM SPSS Statistics. Throughout this book, we refer to it simply as SPSS. Outside this book, that short-cut can be risky because other related products are also called SPSS — notably, IBM SPSS Modeler, which, though powerful and part of the same brand, is not the subject of this book.

About the Fourth Edition

The authors wrote this fourth edition using IBM SPSS Statistics version 27. The technical editor checked all tasks against his concurrent version of SPSS Statistics Subscription and the instructions were consistent on both platforms. The third edition of this book was written using IBM SPSS Statistics version 23. Even if you use a version of SPSS prior to version 27, we urge you to use this fourth edition because its more than 100 pages of new content better prepare new users and meet the needs of intermediate users, as well as cover statistical analytic procedures to a much greater extent.

Foolish Assumptions

This book is for anyone new to SPSS. No prior knowledge of statistics is needed or even expected. In specific terms, we made a few assumptions about you, the reader of this book:

- » You may be a student who isn't majoring in statistics but has been instructed to use SPSS by one of your professors.
- » You may have an interest in statistical analysis but are new to using SPSS.
- » You may work in an office setting and have been asked to use SPSS to analyze some data.

For most people who generate statistics, the complexity of using the software becomes an obstacle. Our purpose in writing this book is to show you how to move that obstacle out of the way with minimum effort.

Icons Used in This Book

Throughout this book, we use icons in the margins to grab your attention. Here's what those icons mean:



REMEMBER

You should keep this information in mind. It's important to what you're doing.



TECHNICAL
STUFF

This icon highlights unnecessarily geeky information, but we had to include it to complete the thought. You can skip anything marked with this icon unless the text makes you curious.



TIP

This icon highlights a point that can save you time and effort.



WARNING

Anything marked with this icon offers information about something that can sneak up and bite you.

Beyond the Book

In addition to the material in the print or e-book you're reading right now, this product also comes with some access-anywhere goodies on the web. Check out the free cheat sheet at www.dummies.com/cheatsheet/spss for information on variable levels of measurement, commonly used procedures in the Analyze menu, and possible conclusions that you can reach after conducting a statistical test. Finally, for the examples, you can download the data files that don't come with SPSS at www.dummies.com/go/spss or keithmccormick.com/SSFD4E.

Where to Go from Here

If you're new to SPSS, we recommend starting out by reading Part 1, so you understand what SPSS is. If you haven't already gained access to a copy of SPSS, check out Chapter 2. Chapter 3 introduces the software and goes through a typical session using SPSS. Read the stuff in Chapter 4 about defining variables and declaring metadata — it all makes sense once you get the hang of it, but the process seems kind of screwy until you see how it works. The material in Chapter 4 covers areas

in which some new users of SPSS go astray, so read it carefully. And from there, use the table of contents and index to find the things you want to do!

Two chapters are worthy of special mention. Chapter 22 helps you navigate statistical techniques covered in Parts 5 and 6 and beyond. If you know the name of the statistical procedure you want to use, you should be okay without Chapter 22. But what if you don't know the name of the statistical procedure? How are you supposed to look it up? This is when you should try Chapter 22.

Sometimes things seem amiss, but you simply don't know what's wrong. Chapter 30 addresses this issue. How are you supposed to find out how to fix an issue when SPSS isn't working as it's supposed to? In this case, check out Chapter 30, which covers the most common reasons why SPSS acts strangely.

If you have a question, you can reach Jesus Salcedo at jesussalcedo@yahoo.com and Keith McCormick at keithmc123@keithmccormick.com. We will post information about updates, joining our mailing list, and the latest SPSS news at keithmccormick.com/SSFD4E.

1

Getting Started with SPSS

IN THIS PART . . .

Find out what SPSS is all about.

Choose the right SPSS.

Get a feel for SPSS with a simple first example session.

- » Considering the quality of your data
- » Communicating with SPSS
- » Seeing how SPSS works
- » Finding help when you're stuck

Chapter 1

Introducing SPSS

A *statistic* is a number, but it's a special kind of number. A statistic is a measurement of some sort. It's fundamentally a count of something — occurrences, speed, amount, or whatever. A statistic is calculated using a sample. In a sense, a sample is the keyhole you have to peer through to see the population, which is what you're trying to understand. The value at the population level — the average height of an American male, for instance — is called a *parameter*. Unless you've got all the data there is, and you've collected a census of the population, you have to make do with the data in your sample. The job of SPSS is to calculate. Your job is to provide a good sample. Together you try to understand the population even though all you have is a sample.

In this chapter, we discuss the importance of having accurate, reliable data, and some of the implications when this is not the case. We talk also about how best to organize your data in SPSS and the different kinds of files that SPSS creates. We take a trip down memory lane and discuss the origins of SPSS so you can understand all of its many names. We discuss what can be done in the program and the different ways of communicating with the software. Finally, we spend some time discussing different ways in which you can get help when navigating SPSS.

SPSS's Job, Our Job, and Your Job

It's important to have appropriate expectations. In this section, we discuss the various roles that all parties play with regard to learning to use SPSS.

SPSS's job

After you give SPSS data and instructions, it will perform the calculations for you. The data and the *metadata* — the information about the data — have to be correct. (We have a lot to say about metadata.) The instructions have to be correct as well. Correct data processed with the wrong technique won't give you the results you need.

SPSS won't make math errors. That's not the kind of errors computers make. They always do *exactly* what we tell them to do, and sometimes that's the problem. SPSS's job is to take data that has been declared correctly and produce statistical results in the form of tables and charts that allow you to draw conclusions about your data — if you know how to interpret those results.

Our job

Your authors, Jesus and Keith, have had so many hours using SPSS that we've lost count. You may have heard of the somewhat controversial *10,000 hour rule*, which states that you need that many hours of "deliberate practice" to truly master a complex subject. Well, you'll be pleased to know that Jesus and Keith each have more than 10,000 classroom hours teaching SPSS in addition to decades of using it on our own projects.

As the authors, we are primarily responsible for the following:

- » **Making the book easy to read:** We know where to focus your attention when you're getting started because we've helped thousands of SPSS beginners get started.
- » **Walking you through how to set up everything properly:** Parts 1 and 2 focus on getting started with SPSS as well as how to define data files, variables, and their attributes.
- » **Acclimating you regarding how SPSS works:** Getting you familiar with the software is our number one focus and mission. We hope to get you to a point where you stop worrying about the software so you can concentrate on your analysis.
- » **Explaining how to tell SPSS to do basic tasks:** This edition has a greatly expanded section in Parts 5 and 6 on understanding the basic theory, choosing the right technique, and interpreting results. We realize that if you're confused about SPSS, it may be because you're confused about statistics. However, we spend the rest of the book helping you understand the software.
- » **Starting you on your SPSS Statistics journey:** You won't master statistics when you reach the end of this book, but you will be much more comfortable with SPSS software. No book of this length could cover all statistical techniques, but we selected the most important procedures for new users of SPSS.

Your job

Your number one job is to relax. You might be reading this book because you're up against a deadline or something in SPSS is causing you stress. However, if you take a little time now to understand how to work in SPSS efficiently, it will pay off in the long run. You have some other responsibilities:

- » **Know your data.** SPSS can't put your data into context — only you can. SPSS will trust everything you give it. It will never second-guess the data you give it, and it will never ask you about it. Only you can ensure that the data is trustworthy.
- » **Declare your data and set it up properly.** Declaring and setting up your data is a critical responsibility that we cover thoroughly, especially in Chapters 3 and 4. Setting up data is not just about declaring the metadata but also about other data management tasks that we cover in Part 3.
- » **Choose the correct statistical technique.** The toughest task for you may be to choose the correct statistical technique. Dozens of techniques are available, but we know which ones are the most important to learn. We've given you a life preserver. If you feel stuck and don't know which chapter to refer to, check out Chapter 22, where we provide an overview for analyzing data and a larger context of where in the research process various statistical techniques are typically used.
- » **Know how to interpret the results.** Finally, Parts 5 and 6 have many examples of how to interpret statistical output.

Garbage In, Garbage Out: Recognizing the Importance of Good Data

SPSS doesn't warn you when there is something wrong with your sample. Its job is to work on the data you give it. If what you give SPSS is incomplete or biased, or if there is data that doesn't belong in there, the resulting calculations won't reflect the population very well. Not much in the SPSS output will signal to anyone that there is a problem. So, if you're not careful, you can conclude just about anything from your data and your calculations.

Consider the data in Table 1-1. What if you calculated the survival rate of *Titanic* passengers based on this small sample? What if you calculated what fraction of the passengers were in each class of service? You can easily see that you'd be in real trouble.

TABLE 1-1**Sample of *Titanic* Passengers**

Survived or Died	Class	Name	Sex	Age	Fare Paid	Cabin	Embarkation
Died	1	Andrews, Mr. Thomas, Jr.	Male	39	0.00	A36	Southampton
Died	1	Parr, Mr. William Henry Marsh	Male		0.00		Southampton
Died	1	Fry, Mr. Richard	Male		0.00	B102	Southampton
Died	1	Harrison, Mr. William	Male	40	0.00	B94	Southampton
Died	1	Reuchlin, Mr. John George	Male	38	0.00		Southampton
Died	2	Parkes, Mr. Francis "Frank"	Male		0.00		Southampton
Died	2	Cunningham, Mr. Alfred Fleming	Male		0.00		Southampton
Died	2	Campbell, Mr. William	Male		0.00		Southampton
Died	2	Frost, Mr. Anthony Wood "Archie"	Male		0.00		Southampton
Died	2	Knight, Mr. Robert J.	Male		0.00		Southampton
Died	2	Watson, Mr. Ennis Hastings	Male		0.00		Southampton
Died	3	Leonard, Mr. Lionel	Male	36	0.00		Southampton
Died	3	Tornquist, Mr. William Henry	Male	25	0.00		Southampton
Died	3	Johnson, Mr. William Cahoone, Jr.	Male	19	0.00		Southampton
Died	3	Johnson, Mr. Alfred	Male	49	0.00		Southampton

However, consider this: Would you be tempted to drop these cases from your analysis because their fare information appears to be missing? What if fare information were provided for all the other passengers? You might drop the cases in Table 1-1 but use everyone else. You'd be dropping only a handful of passengers out of hundreds, so that would be okay, right? The answer is no, it would not be okay. As it turns out, there is a good reason that each of these passengers didn't pay a fare (for example, Mr. Thomas Andrews, Jr., designed the ship), and if this was your data, your job would be to know that.

Sampling is a big topic, but here's the quick version:

- » The data points in your sample should be drawn at random from the population.
- » There should be enough data points.
- » You should be able to justify the removal of any data points.

THE ORIGIN OF SPSS

In 2018, IBM SPSS Statistics turned 50. That makes it older than Windows and older than the first Apple computer, so in the early days SPSS was run on mainframe computers using punch cards.

At Stanford University in the late 1960s, Norman H. Nie, C. Hadlai (Tex) Hull, and Dale H. Bent developed the original software system named Statistical Package for the Social Sciences (SPSS). They needed to analyze a large volume of social science data, so they wrote software to do it. The software package caught on with other folks at universities, and, consistent with the open-source tradition of the day, the software spread through universities across the country.

The three men produced a manual in the 1970s, and the software's popularity took off. A version of SPSS existed for each of the different kinds of mainframe computers in existence at the time. Its popularity spread from universities into the public sector, and it began to leak into the private sector as well.

In the 1980s, a version of the software was moved to the personal computer. In 1992, Jack Noonan became CEO of SPSS, Inc. (replacing Nie) and a period of acquisition of smaller software companies began. Many of those products are still part of the SPSS family, such as IBM SPSS AMOS (for structural equation modeling) and IBM SPSS Modeler (originally called Clementine). In 2009, SPSS, Inc. was acquired by IBM, and the name of the product became IBM SPSS Statistics to differentiate it from the other products.

The official name of the software today is still IBM SPSS Statistics, and it's available in several formats and versions. We discuss these different options in Chapter 2.



REMEMBER

This book is not about the accuracy, correctness, or completeness of the input data. Your data is up to you. This book shows you how to take the numbers you already have, put them into SPSS, crunch them, and display the results in a way that makes sense. Gathering valid data and figuring out which cases to use is up to you.



WARNING

Your data is your most valuable possession. If you're the only one in the world with your data, be sure to back it up before you start working with it. Make sure you have multiple copies, ideally with one copy in the cloud. At key milestones in your analysis and data modifications, remember to save it again. The last thing you want is to lose your data.

Talking to SPSS: Can You Hear Me Now?

More than one way exists for you to command SPSS to do your bidding. You can use any of three approaches to perform any of the SPSS functions, and we cover them all in this section. The method you should choose depends not only on which interface you prefer, but also (to an extent) on the task you want performed.

The graphical user interface

SPSS has a window interface. You can issue commands by using the mouse to make menu selections that cause dialog boxes to appear. This is a fill-in-the-blanks approach to statistical analysis that guides you through the process of making choices and selecting values. The advantage of the graphical user interface (GUI) approach is that, at each step, SPSS makes sure you enter everything necessary before you can proceed to the next step. This interface is preferred for those just starting out — and if you don't go into depth with SPSS, this may be the only interface you ever use.

Syntax

Syntax is the internal language used to command actions from SPSS. It's the command syntax of SPSS (hence, its name). Syntax is often referred to as the "command language." You can use the Syntax command language to enter instructions into SPSS and have it do anything it's capable of doing. In fact, when you select from menus and dialog boxes to command SPSS, you're actually generating Syntax commands internally that do your bidding. In other words, the GUI is nothing more than the front end of a Syntax command-writing utility.

Writing (and saving) command-language programs is a good way to create processes that you expect to repeat. You can even grab a copy of the Syntax commands generated from the menu and save them to be repeated later.

Programmability

Programmability refers to the myriad ways of customizing SPSS with extensions. These *extensions* are new capabilities that the user community adds to SPSS using the programming languages Python and R. Learning to write these powerful new features is beyond the scope of this book, but you should know that they exist. (SPSS has an entire menu called Extensions.) When you allow SPSS to do so, a number of these extensions are installed during installation.

How SPSS works

The developers of SPSS have made every effort to make the software easy to use. SPSS prevents you from making mistakes or even forgetting something. That's not to say it's impossible to do something wrong in SPSS, but SPSS software works hard to keep you from running into the ditch. To foul things up, you almost have to work at figuring out a way of doing something wrong.

You always begin by defining a set of *variables*; then you enter data for the variables to create a number of *cases*. For example, if you're doing an analysis of automobiles, each car in your study would be a case. The variables that define the cases could be things such as the year of manufacture, horsepower, and cubic inches of displacement. Each car in the study is defined as a single case, and each case is defined as a set of values assigned to the collection of variables. Every case has a value for each variable. (Well, you *can* have a missing value, but that's a special situation described later.)

There are different types of variables. These types describe how the data is *stored* — for example, as letters (strings), as numbers, as dates, or as currency (see Chapter 4 for more information on data types). Each variable is defined as containing a certain kind of number, so you also have to define the variable's level of measurement. For example, a *scale* variable is a numeric measurement, such as weight or miles per gallon. A *categorical* variable contains values that define a category; for example, a variable named `gender` could be a categorical variable defined to contain only values 1 for female and 2 for male. Things that make sense for one type of variable don't necessarily make sense for another. For example, it makes sense to calculate the average miles per gallon, but not the average gender.

After your data is entered into SPSS — your cases are all defined by values stored in the variables — you can easily run an analysis. You've already finished the hard part. Running an analysis on the data is simple compared to entering the data. To run an analysis, you select the analysis you want to run from the menu, select the appropriate variables, and click OK. SPSS reads through all your cases, performs the analysis, and presents you with the output as tables or graphs. Of course, you have to know which analysis to choose. For that information, see Parts 5 and 6.

You can instruct SPSS to draw graphs and charts directly from your data the same way you instruct it to do an analysis. You select the desired graph from the menu, assign variables to it, and click OK.



TIP

When you're preparing SPSS to run an analysis or draw a graph, the OK button is unavailable until you've made all the choices necessary to produce output. Not only does SPSS require that you select a sufficient number of variables to produce output, but it also requires you to choose the right kinds of variables. If a categorical variable is required for a certain slot, SPSS won't allow you to choose any other kind of variable. Whether the output makes sense is up to you and your data, but SPSS makes sure that the choices you make can be used to produce some kind of result.

NUMBERS NOT WORDS

SPSS works best with numbers. Whenever possible, try to have your SPSS data in the form of numbers. If you give SPSS names and descriptions, it'll seem like they're being processed by SPSS, but that's because each name has been assigned a number. (Sneaky.) That's why survey questions are written like this:

- How do you feel about rhubarb? Select one answer:
 - A. I love it!
 - B. It's okay.
 - C. I can take it or leave it.
 - D. I don't care for it.
 - E. I hate it!

A number is assigned to each of the possible answers, and these numbers are fed through the statistical process. SPSS uses the numbers — not the words — so be careful about keeping all your words and numbers straight. We cover this subject in some detail in Chapter 4.

Remember: Keep accurate records describing your data, how you got the data, and what it means. SPSS can do all the calculations for you, but only you can decipher what it means. In *The Hitchhiker's Guide to the Galaxy*, a computer the size of a planet crunched on a problem for generations and finally came out with the answer, 42. But the people tending the machine had no idea what the answer meant because they didn't remember the question. They hadn't kept track of their input. You must keep careful track of your data or you may later discover, for example, that what you've interpreted to be a simple increase is actually an increase in your rate of decrease. Oops!

MAKING SENSE OF ALL THOSE SPSS FILES

Input data and statistics are stored in files — different kinds of files. Some files contain numbers and definitions of numbers. Some files contain graphics. Some files contain both. Data files are easy to spot because they end with the .sav extension. Output files end with the .spv extension. Command Syntax files, with the optional programming language commands, end with .sps.

The examples in this book require the use of files that contain data configured to demonstrate capabilities of SPSS. Most of the files are in the same directory you used to install SPSS (installing SPSS also installs a number of data files ready to be loaded into SPSS and used for analysis). A few of the files used in the examples can be downloaded from the book's companion website (www.dummies.com/go/spss).

All output from SPSS goes to the same place — a window named SPSS Statistics Viewer. This window displays the results of whatever you've done. After you've produced output, if you perform some action that produces more output, the new output is displayed in the same window. And almost anything you do produces output. Of course, you need to know how to interpret the output — SPSS will help you, and so does this book.

Getting Help When You Need It

You're not alone. Some immediate help comes directly from the SPSS software package. More help can be found online. If you find yourself stumped, you can look for help in several places:



TIP

» **Topics:** Choosing Help ⇄ Topics from the main window of the SPSS application is your gateway to immediate help. The help is somewhat terse, but it usually provides exactly the information you need. The information is in one large Help document, presented one page at a time. Choose Contents to select a heading from an extensive table of contents, choose Index to search for a heading by entering its name, or choose Search to enter a search string inside the body of the Help text.

In the Help directory, the titles in all uppercase are descriptions of Syntax language commands.

» **SPSS Support:** Choose Help ⇄ Support to open a browser window for the support page at IBM. This area is primarily to report potential bugs or to check

if anyone else has encountered the same bug. It's not the best option if you're struggling with a task on the first try.

- » **SPSS Support Forums:** Choose Help ⇨ SPSS Forums to open a browser showing the various support forums. IBM is putting a lot of resources into SPSS communities, which might have more activity over time than these forums.
- » **PDF Documentation:** Choose Help ⇨ Documentation in PDF Format if you want to access the many user's guides for SPSS. This resource is online, but you can download them all to a folder on your machine if you want offline access to them.
- » **Command Syntax Reference:** Choose Help ⇨ Command Syntax Reference to display more than 2,000 pages of references to the Syntax language in your PDF viewer. The regular help topics, mentioned previously, provide a brief overview of each topic, but this document is more detailed.
- » **Compatibility Report Tool:** Choose Help ⇨ Compatibility Report Tool to answer a series of queries online to determine the compatibility of your software and hardware. If you're having trouble getting SPSS to install, access this information at www.ibm.com/software/reports/compatibility/clarity/index.html.
- » **SPSS Statistics Community:** Choose Help ⇨ IBM SPSS Predictive Analytics Community to visit a huge collection of IBM blogs and forums for every need. It will take a little time to get registered and settled in, but it's designed to be your free, go-to resource for the latest news and a chance to interact with other users. Be sure to sign up for the SPSS Stats group, in the IBM Data Science community. Hundreds of thousands of people are in this community, so it should be your first stop, before the support forums.