

doing data analysis

with SPSS[®] 10.0

Carver & Nash

Doing Data Analysis with SPSS 10.0

Robert H. Carver

Jane Gradwohl Nash

Stonehill College



Sponsoring Editor: *Seema Atwal*
Marketing Team: *Samantha Cabaluna, Tom Ziolkowski*
Editorial Assistant: *Emily Davidson*
Production Coordinator: *Kelsey McGee*
Production Service: *Matrix Productions/
Merrill Peterson*

Manuscript Editor: *Kathy Behler*
Permissions Editor: *Mary Kay Hancharick*
Cover Design: *Laurie Albrecht*
Cover Photo: *PhotoDisc*
Print Buyer: *Vena Dyer*
Cover Printing, Printing and Binding: *Webcom, Ltd.*

COPYRIGHT © 2000 by Brooks/Cole
Duxbury is an imprint of Brooks/Cole, a division of Thomson Learning
The Thomson Learning logo is a trademark used herein under license.

For more information about this or any other Duxbury product, contact:

DUXBURY

511 Forest Lodge Road

Pacific Grove, CA 93950 USA

www.duxbury.com

1-800-423-0563 Thomson Learning Academic Resource Center)

All rights reserved. No part of this work may be reproduced, transcribed or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, Web distribution, or information storage and/or retrieval systems—without the prior written permission of the publisher.

For permission to use material from this work, contact us by

Web: www.thomsonrights.com

fax: 1-800-730-2215

phone: 1-800-730-2214

Printed in Canada

10 9 8 7 6 5 4

Library of Congress Cataloging-in-Publication Data

Carver, Robert H.

Doing Data Analysis with SPSS 10.0/Robert H. Carver, Jane Gradwohl Nash.

p. cm.

ISBN 0 534-37475-1

1. SPSS (Computer file) 2. Social sciences—Statistical methods

Computer programs I. Nash, Jane Gradwohl. II. Title.

HA32.C37 2000

00-036091

519.5'0285'5369—dc2100



Preface

The Changing Environment of Statistics Education

In the past decade or so, educators have come to reconsider the best approach to teaching and learning in applied statistics courses. With the widespread availability of personal computers, advances in statistical software, and the near-universal application of quantitative methods in many professions, courses now emphasize statistical reasoning more than computational skill development. Questions of *how* have given way to more challenging questions of *why*, *when*, and *what*?

Simultaneously, undergraduates are increasingly comfortable with software, expecting to use computers in their work. Colleges are seeking ways to integrate information technology efficiently into coursework. The introductory statistics course is an ideal place to augment traditional out-of-class assignments with structured computer exercises.

The goal of this book is to supplement an introductory undergraduate statistics course with a comprehensive set of self-paced exercises. Students can work independently, learning the software skills outside of class, while coming to understand the underlying statistical concepts and techniques. Instructors can teach statistics and statistical reasoning, rather than algebra or software.

The Approach of This Book

The book reflects the changes described above in several ways. First, and most obviously, it provides some training in the use of a powerful software package to relieve students of computational drudgery.

Second, each session is designed to address a statistical issue or need, rather than to feature a particular command or menu in the software. Third, nearly all of the datasets in the book are real, reflecting a variety of disciplines. Fourth, the sessions follow a traditional sequence, making the book compatible with many texts. Finally, as each session leads the student through the techniques, it also includes thought-provoking questions and challenges, engaging the student in the processes of statistical reasoning. In designing the lab exercises, we kept four ideas in mind:

- *Statistical reasoning, not computation, is the goal of the course.* This manual asks students questions throughout, balancing software instruction with reflection on the meaning of results.
- *Students arrive in the course ready to learn statistical reasoning.* They need not slog all the way through descriptive techniques before encountering the concept of inference. The exercises invite students to think about inferences from the start, and the questions grow in sophistication as students master new material.
- *Exploration of real data is preferable to artificial datasets.* With the exception of the famous Anscombe regression dataset and a few simulations, all of the datasets are real. Some are very old and some are quite current, and they cover a wide range of substantive areas.
- *Statistical topics, rather than software features, should drive the design of each lab session.* Each lab session features several SPSS functions selected for their relevance to the statistical concept under consideration.

This book provides a rigorous but limited introduction to the software. The SPSS Base 10.0 system is rich in features and options; this book makes no attempt to “cover” the entire package. Instead, the level of coverage is commensurate with an introductory course. There may be many ways to perform a given task in SPSS; generally, we show one way. This book provides a “foot in the door.” Interested students and other users can explore the software possibilities via the extensive Help system or other standard SPSS documentation.

Using This Book

We presume that this book is being used as a supplementary text in an introductory-level statistics course. If your courses are like ours

(one in a psychology department, the other in a business department), class time is a scarce resource. Adding new material is always a balancing act. As such, supplementary readings and assignments must be carefully integrated. We suggest that instructors use the sessions in this book in four different ways, tailoring the approach throughout the term to meet the needs of the students and course.

- *In-class activity:* Part or all of some sessions might best be done together in class, with each student at a computer. The instructor can comment on particular points and can roam to offer assistance. This may be especially effective in the earliest sessions.
- *Stand-alone assignments:* In conjunction with a topic covered in the principal text, sessions can be assigned as independent out-of-class work, along with selected Moving On... questions. This is our most frequently-used approach. Students independently learn the software, re-enforce the statistical concepts, and come to class with questions about any difficulties they encountered in the lab session.
- *Preparation for text-based case or problem:* An instructor may wish to use a textbook case for a major assignment. The relevant session may prepare the class with the software skills needed to complete the case.
- *Independent projects:* Sessions may be assigned to prepare students to undertake an independent analysis project designed by the instructor. Many of the data files provided with the book contain additional variables that are never used within sessions. These variables may form the basis for original analyses or explorations.

Solutions are available to instructors for all Moving On... and bold-faced questions. Instructors should consult their Duxbury sales representatives for details.

The Data Files

As previously noted, each of the data files provided with this book contains real data, much of it downloaded from public sites on the World Wide Web. You can download all files from the Duxbury Press web site. Appendix A describes each file and its source, and provides detailed definitions of each variable.

The data files were chosen to represent a variety of interests and fields, and to illustrate specific statistical concepts or techniques. No

doubt, each instructor will have some favorite datasets that can be used with these exercises. Most textbooks provide datasets as well. For some tips on converting other datasets for use with SPSS, see Appendix B.

Note on Software Versions

The examples in this manual are based on SPSS Base 10.0, running under Windows 95, Windows 98, or Windows NT. Users of earlier Windows versions or the Student version will notice only minor differences with the figures and instructions in this book, and in a few instances, will need to take an alternate approach. Adopters using SPSS Base 9.0 should assign Appendix C in lieu of Session 1.

To the Student

This book has two goals: to help you understand the concepts and techniques of statistical analysis, and to teach you how to use one particular tool—SPSS—to perform such analysis. It can supplement but not replace your primary textbook or your classroom time. To get the maximum benefit from the book, you should take your time and work carefully. Read through a session before you sit down at the computer. Each session should require no more than about 30 minutes of computer time; there's little need to rush through them.

We have included dialog box images for each new command. In those instances where you must drag a variable name into a box, you will notice that the name sometimes overhangs the box. Upon release of the mouse button, the name is truncated.

You'll often see questions interspersed through the computer instructions. These are intended to shift your focus from "getting answers" to thinking about what the answers mean, whether they make sense, whether they surprise or puzzle you, or how they relate to what you have been doing in class. Attend to these questions, even when you aren't sure of their purpose.

You may also notice that we have sometimes placed our names on a graph. We do this intermittently to call your attention to the practice; you should always place your name on the graphs you create.


Each lab ends with a section called ***Moving On....*** You should also respond to the numbered questions in that section, as assigned by your instructor. Questions in the Moving On... sections are designed to challenge you. Sometimes, it is quite obvious how to proceed with your analysis; sometimes, you will need to think a bit before you issue your first command. The goal is to get you to engage in statistical thinking.

integrating what you have learned throughout your course. There is much more to doing data analysis than “getting the answer,” and these questions provide an opportunity to do realistic analysis.

As noted earlier, SPSS is a large and very powerful software package, with many capabilities. Many of the features of the program are beyond the scope of an introductory course, and do not figure in these exercises. However, if you are curious or adventurous, you should explore the menus and Help system. You may find a quicker, more intuitive, or more interesting way to approach a problem.

Typographical Conventions

Throughout this manual, certain symbols and typefaces are used consistently. They are as follows:

 **Menu > Sub-menu > Command** The mouse icon indicates an action you take at the computer, using the mouse or keyboard. The bold type lists menu selections for you to make.

Dialog box headings are in this typeface.

Dialog box choices, variable names, and items you should type appear in this typeface.

File names (e.g., *Colleges*) appear in this typeface.



A box like this contains an instruction requiring special care or information about something that may work differently on your computer system.

Bold italics in the text indicate a question that you should answer as you write up your experiences in the lab.

Acknowledgments

Like most authors, we owe many debts of gratitude for this book. This project enjoyed the support of Stonehill College through the annual Summer Grants and the Stonehill Undergraduate Research Experience (SURE) programs. As the SURE scholar in the preparation of the book, Jason Boyd contributed in myriad ways, consistently doing reliable, thoughtful, and excellent work. He tested every session, prepared instructors' solutions, researched datasets, critiqued sessions from a student perspective, and tied up loose ends. His contributions and collegiality were invaluable.

Many colleagues and students suggested or provided datasets. Student contributors were Jennifer Axon, Stephanie Duggan, Debra Elliott, Tara O'Brien, Erin Ruell, and Benjamin White. A big thank you goes out to our students in Introduction to Statistics and Quantitative Analysis for Business for pilot-testing many of the sessions and for providing useful feedback about them.

We thank our Stonehill colleagues Ken Branco, Lincoln Craton, Roger Denome, Jim Kenneally, and Bonnie Klentz for suggesting or sharing data, and colleagues from other institutions who supported our work: Chris France, Roger Johnson, Stephen Nissenbaum, Mark Popovksy, and Alan Reifman. Thanks also to the many individuals and organizations granting permission to use published data for these sessions; they are all identified in Appendix A.

At Duxbury Press, we enjoyed the guidance and encouragement of Curt Hinrichs, Carolyn Crockett, Sarah Kaminskis, and Seema Atwal. Thanks also go to Paul Baum at California State University, Northridge and to Dennis Jowaisas at Oklahoma City University, two reviewers whose constructive suggestions have improved the quality of this book.



Finally, we thank our families.

I want to thank my husband, Justin Nash, for his unwavering support of my professional work, and my daughters, Hanna Gradwohl Nash and Sara Gradwohl Nash, for providing an enjoyable distraction from this project.

JGN

The Carver home team has been fabulous, as always. To Donna, my partner and counsel; to Sam and Ben, my cheering section and assistants. Thanks for the time, space, and encouragement. Sometimes it *does* help to hear, "Dad, why are you writing another book?"

RHC

About the Authors

Robert H. Carver is Professor of Business Administration at Stonehill College in Easton, Massachusetts where his teaching has been recognized with the College's annual Excellence in Teaching award. In addition to Business Statistics, he teaches courses in information systems as well as business and society. He holds an A.B. from Amherst College and a Ph.D. in Public Policy from the University of Michigan. He is the author of *Doing Data Analysis with Minitab 12* (Duxbury Press), and his work has appeared in *Publius*, *The Journal of Statistics Education*, *PS: Political Science & Politics*, *Public Administration Review*, *Public Productivity Review*, and *The Journal of Consumer Marketing*.

Jane Gradwohl Nash is Associate Professor of Psychology at Stonehill College. She earned her B.A. from Grinnell College and her Ph.D. from Ohio University. She enjoys teaching courses in the areas of statistics, cognitive psychology, and developmental psychology. Her research interests are in the area of knowledge structure and knowledge change (learning). She is the author of articles that have appeared in the *Journal of Educational Psychology*, *Organizational Behavior and Human Decision Processes*, *Journal of Chemical Education*, *Research in the Teaching of English*, and *Written Communication*.

Contents

Session 1.	A First Look at SPSS 10.0	1
Objectives	1	
Launching SPSS	1	
Entering Data into the Data Editor	3	
Saving a Data File	7	
Creating a Bar Chart	7	
Saving an Output File	11	
Getting Help	12	
Printing in SPSS	12	
Quitting SPSS	12	
Session 2.	Tables and Graphs for One Variable	13
Objectives	13	
Opening a Data File	13	
Exploring the Data	14	
Creating a Histogram	16	
Frequency Distributions	19	
Another Bar Chart	21	
Printing Session Output	21	
Moving On...	22	
Session 3.	Tables and Graphs for Two Variables	25
Objectives	25	
Cross-Tabulating Data	25	
Editing a Recent Dialog	27	
More on Bar Charts	27	
Comparing Two Distributions	30	

Scatterplots to Detect Relationships	31
Moving On...	33
Session 4. One-Variable Descriptive Statistics	37
Objectives	37
Computing One Summary Measure for a Variable	37
Computing Additional Summary Measures	41
A Box-and-Whiskers Plot	44
Standardizing a Variable	45
Moving On...	46
Session 5. Two-Variable Descriptive Statistics	49
Objectives	49
Comparing Dispersion with the Coefficient of Variation	49
Descriptive Measures for Subsamples	51
Measures of Association: Covariance and Correlation	52
Moving On...	54
Session 6. Elementary Probability	59
Objectives	59
Simulation	59
A Classical Example	59
Observed Relative Frequency as Probability	61
Handling Alphanumeric Data	63
Moving On...	66
Session 7. Discrete Probability Distributions	69
Objectives	69
An Empirical Discrete Distribution	69
Graphing a Distribution	71
A Theoretical Distribution: The Binomial	72
Another Theoretical Distribution: The Poisson	74
Moving On...	75
Session 8. Probability Density Functions	79
Objectives	79
Continuous Random Variables	79
Generating Normal Distributions	80
Finding Areas under a Normal Curve	83
Normal Curves as Models	85
Moving On...	87

Session 9. Sampling Distributions	91
Objectives	91
What Is a Sampling Distribution?	91
Sampling from a Normal Population	92
Central Limit Theorem	95
Sampling Distribution of the Proportion	97
Moving On...	98
Session 10. Confidence Intervals	101
Objectives	101
The Concept of a Confidence Interval	101
Effect of Confidence Coefficient	104
Large Samples from a Non-normal (Known) Population	104
Dealing with Real Data	105
Small Samples from a Normal Population	106
Moving On...	108
Session 11. One-Sample Hypothesis Tests	111
Objectives	111
The Logic of Hypothesis Testing	111
An Artificial Example	112
A More Realistic Case: We Don't Know μ or σ	115
A Small-Sample Example	117
Moving On...	120
Session 12. Two-Sample Hypothesis Tests	123
Objectives	123
Working with Two Samples	123
Paired vs. Independent Samples	128
Moving On...	130
Session 13. Analysis of Variance (I)	135
Objectives	135
Comparing Three or More Means	135
One-Factor Independent Measures ANOVA	136
Where Are the Differences?	140
One-Factor Repeated Measures ANOVA	142
Where Are the Differences?	147
Moving On...	147

Session 14. Analysis of Variance (II)	151
Objectives	151
Two-Factor Independent Measures ANOVA	151
Another Example	157
One Last Note	159
Moving On...	160
Session 15. Linear Regression (I)	163
Objectives	163
Linear Relationships	163
Another Example	168
Statistical Inferences in Linear Regression	169
An Example of a Questionable Relationship	170
An Estimation Application	171
A Classic Example	172
Moving On...	173
Session 16. Linear Regression (II)	177
Objectives	177
Assumptions for Least Squares Regression	177
Examining Residuals to Check Assumptions	178
A Time Series Example	183
Issues in Forecasting and Prediction	185
A Caveat about "Mindless" Regression	188
Moving On...	189
Session 17. Multiple Regression	193
Objectives	193
Going Beyond a Single Explanatory Variable	193
Significance Testing and Goodness of Fit	199
Residual Analysis	200
Adding More Variables	200
Another Example	201
Working with Qualitative Variables	202
A New Concern	204
Moving On...	205
Session 18. Nonlinear Models	209
Objectives	209
When Relationships Are Not Linear	209

A Simple Example	210
Some Common Transformations	211
Another Quadratic Model	213
A Log-Linear Model	217
Adding More Variables	218
Moving On...	219
Session 19. Basic Forecasting Techniques	223
Objectives	223
Detecting Patterns over Time	223
Some Illustrative Examples	224
Forecasting Using Moving Averages	226
Forecasting Using Trend Analysis	229
Another Example	232
Moving On...	232
Session 20. Chi-Square Tests	235
Objectives	235
Qualitative vs. Quantitative Data	235
Chi-Square Goodness-of-Fit Test	235
Chi-Square Test of Independence	239
Another Example	242
Moving On...	243
Session 21. Nonparametric Tests	247
Objectives	247
Nonparametric Methods	247
Mann-Whitney U Test	248
Wilcoxon Signed Ranks Test	250
Kruskal-Wallis H Test	252
Spearman's Rank Order Correlation	255
Moving On...	256
Session 22. Tools for Quality	259
Objectives	259
Processes and Variation	259
Charting a Process Mean	260
Charting a Process Range	263
Another Way to Organize Data	264
Charting a Process Proportion	266
Pareto Charts	268
Moving On...	270

Appendix A:	Dataset Descriptions	273
Appendix B:	Working with Files	309
	Objectives	309
	Data Files	309
	Viewer Document Files	310
	Converting Other Data Files into SPSS Data Files	311
Appendix C:	A First Look at SPSS 9.0	315
	Objectives	315
	Launching SPSS	315
	Entering Data into the Data Editor	317
	Saving a Data File	321
	Creating a Bar Chart	321
	Saving an Output File	325
	Getting Help	326
	Printing in SPSS	326
	Quitting SPSS	326
Index		327

Session 1

A First Look at SPSS 10.0

Objectives

In this session, you will learn to do the following:


- Launch and exit SPSS
- Enter quantitative and qualitative data in a data file
- Create and print a graph
- Get Help
- Save your work to a disk

Launching SPSS

Before starting this session, you should know how to run a program within the Windows 95, 98, or Windows NT operating system. All the instructions in this manual presume basic familiarity with the Windows environment.



Check with your instructor for specific instructions about running Windows 95/98/NT on your system. Your instructor will also tell you where to find SPSS.

Click and hold the left mouse button on the  **Start** button at the lower left of your screen, and drag the cursor to select **Programs**. In the list, locate and choose **SPSS 10.0 for Windows**. Click and release the mouse button to launch the program. Because SPSS is a large program, you may have to wait a few moments before the program is ready for use.

On the next page is an image of the screen you will see when SPSS is ready. First you will see a menu dialog box listing several