





SPSS/PC+ *Studentware*[™]



MARIJA J. NORUŠIS
SPSS INC.



SPSS/PC+ *Studentware*TM

MARIJA J. NORUŠIS

**Rush-Presbyterian-St. Luke's
Medical Center**

**SPSS Inc.
444 N. Michigan Avenue
Chicago, Illinois 60611
312.329.3300**

**SPSS International B.V.
P.O. Box 115
4200 AC Gorinchem
The Netherlands
Tel. +31.1830.36711
Twx.: 21019 (SPSS NL)
Fax +31.1830.35839**

For more information about SPSS/PC+™ Studentware, the SPSS/PC+™ system, and other software produced and distributed by SPSS Inc., please write or call

Marketing Department
SPSS Inc.
444 North Michigan Avenue
Chicago, IL 60611
312/329-3500

In Europe and the Middle East, please write or call

SPSS International B.V.
P.O. Box 115
4200 AC Gorinchem
The Netherlands
+31.1830.36711
Twx: 21019 (SPSS nl)
Fax: +31.1830.35839

SPSS is a registered trademark and the other product names are the trademarks of SPSS Inc. for its proprietary computer software. No material describing such software may be produced or distributed without the written permission of the owners of the trademark and license rights in the software and the copyrights in the published materials.

The SOFTWARE and documentation are provided with RESTRICTED RIGHTS. Use, duplication, or disclosure by the Government is subject to restrictions as set forth in subdivision (c)(1)(ii) of The Rights in Technical Data and Computer Software clause at 52.227-7013. Contractor/manufacturer is SPSS Inc., 444 N. Michigan Avenue, Chicago, IL, 60611.

IBM, IBM PC, IBM PC/XT, IBM PC/AT, and PS/2 are registered trademarks of International Business Machines Corporation.

SPSS/PC+ Studentware™
Copyright © 1988 by SPSS Inc.
All rights reserved.
Printed in the United States of America.

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, or otherwise, without the prior written permission of the publisher.

5 6 7 8 9 0 92 91 90

ISBN 0-918469-73-2

Library of Congress Catalog Card Number: 88-080965

Preface

“The business of a poet is to examine not the individual but the species; to remark general properties and large appearances. He does not number the streaks of the tulip, or describe the different shades of verdure of the forest; he is to exhibit . . . such prominent and striking features as recall the original to every mind.” (Samuel Johnson, *Rasselas*.)

It is not often that anyone compares a statistician to a poet. Yet it is fitting to do so. The statistician, like Johnson’s poet, searches for “general properties and large appearances.” The goal of data analysis is to describe the species based on observations of individuals. The data analyst must identify patterns from thousands of fragments and then speak of the whole. For that is what is of interest.

In this book we consider how to proceed from individual observations to the whole. There are many ways in which the fragments may be assembled, and these can result in different views of the whole. Statistics books differ in what parts of the assembly process they emphasize. This book tries to give students the skills that they need to become informed consumers or producers of statistical information. Therefore this book emphasizes what the statistical process is all about: how to conduct studies, what the results mean, and what can be said about the whole from the pieces.

Since personal computers are routinely used for data analysis today, this book also introduces students both to SPSS/PC+ and to SPSS/PC+ Studentware, a smaller version of SPSS/PC+ intended for student use. *The book is designed for use both with Studentware and with SPSS/PC+.* The commands used are identical in both systems. Chapters 4 and 5 include all the steps needed to use Studentware on a floppy-disk-based PC. Appendix C explains the simpler procedures to use with a hard disk, and Appendix D describes some of the additional features available in the full SPSS/PC+ system. With either SPSS/PC+ or Studentware, students can practice using the tools for analyzing data that the professionals use, and gain experience in analyzing data the way professional researchers do. They need these skills if they pursue graduate degrees, and equally if they enter the working world.

Using this Text

This book is in four parts: preparing data for analysis, describing data, testing hypotheses, and examining relationships. Examples from the NORC General Social Survey are used throughout. Of course, the best

way to learn about anything is to actually do it. That's especially true for data analysis, so each chapter closes with exercises that reinforce and extend the material in three main areas: syntax, statistical concepts, and data analysis. These exercises test understanding of both the mechanics of statistical analysis and the interpretation of the results. The data used in the data analysis exercises, a subset of the General Social Survey data, are available from SPSS Inc. (and are included on the Studentware diskettes). Appendix B contains selected answers to the exercises. Appendix A explains how to correct some of the errors commonly made by users of the SPSS/PC+ system.

Acknowledgments

I wish to thank the members of the SPSS staff who have participated in the preparation of this book. I have benefited from their expertise. Many reviewers have offered helpful comments and suggestions, for which I am grateful. Finally, I wish to thank Professor James A. Schoenberger for encouraging my writing efforts.

Marija J. Norušis

Contents

Part 1

1

INTRODUCTION 2

- What Are Data? 2
 - What Does Data Analysis Involve? 3
 - Why Use Computers? 3
- What Is SPSS/PC+? 3
 - SPSS/PC+ Studentware 4
- About this Book 4
 - Designing a Study 4
 - Describing Data 5
 - Testing Hypotheses 5
 - Describing Relationships 5

2

DESIGNING A STUDY 6

- Asking a Question 7
 - What Information Do You Need? 7
- Defining a Population 8
 - Sampling 9
- Using Surveys 11
 - Conducting Your Own Survey 12
 - Analyzing an Existing Survey 12
- Designing Experiments 13
 - Random Assignment 13
 - "Blind" Experiments 15
- What's Next? 17

3

DESIGNING FORMS FOR STUDIES 20

- The General Social Survey 21
 - Is Life Exciting? 21
- Designing the Form 22
 - Arranging the Form 22
 - Coding the Data 23
 - Tips on Form Design 26
- Collecting the Data 27
- What's Next? 28

4

SETTING UP YOUR DATA FILE 32

- What is a Computer File? 33
- Preparing to Enter Data 34
 - Determining the Data Format 34
 - Fixed Format 34
 - Freefield Format 36

REVIEW, The SPSS/PC+ Text Editor	37
----------------------------------	----

Using REVIEW to Create a Data File	37
------------------------------------	----

Preparing a Data Diskette	38
---------------------------	----

An Example	38
------------	----

REVIEW Commands	40
-----------------	----

Entering the Data	41
-------------------	----

Tips on Entering Data	44
-----------------------	----

More about REVIEW	45
-------------------	----

Other Features	46
----------------	----

Shortcuts	46
-----------	----

What's Next?	47
--------------	----

5

RUNNING SPSS/PC+ 50

Starting Up the System	51
------------------------	----

Making a Backup Copy	51
----------------------	----

The Data Diskette	52
-------------------	----

Beginning Your Session	52
------------------------	----

Quitting	53
----------	----

SPSS/PC+ Commands	53
-------------------	----

The SPSS/PC+ Language	54
-----------------------	----

Getting Help	55
--------------	----

The SPSS/PC+ Session	56
----------------------	----

Identifying the Data	56
----------------------	----

Modifying the Data	57
--------------------	----

Analyzing the Data	57
--------------------	----

Saving Your Data	57
------------------	----

What's Next?	58
--------------	----

6

DEFINING YOUR DATA 60

Cases, Variables, and Values	61
------------------------------	----

Identifying Your Data File	61
----------------------------	----

The DATA LIST Command	62
-----------------------	----

Identifying Your Variables	62
----------------------------	----

Variable Names	62
----------------	----

Naming Variables on a DATA LIST	64
---------------------------------	----

Defining Records	64
------------------	----

String Variables	66
------------------	----

DATA LIST for Freefield Data	67
------------------------------	----

Including Your Data in the Command File	67
---	----

The Active File	68
-----------------	----

Missing Data	68
--------------	----

User-Missing and System-Missing Values	69
--	----

Variable Labels	69
-----------------	----

Value Labels	70
--------------	----

An Example: Defining the General Social Survey Data	71
---	----

Saving and Using System Files	73
-------------------------------	----

Saving a System File	74
----------------------	----

Saving the GSS File	74
---------------------	----

The LIST Command	75
------------------	----

How to Read a System File	76
---------------------------	----

What's Next?	76
--------------	----

7**COUNTING RESPONSES FOR A SINGLE VARIABLE**

80

Counting Frequencies	81
Interpreting a Frequency Table	81
Checking the Data	82
Percentages	83
Valid Percentages	83
Bar Charts	84
Cumulative Percentages	85
Levels of Measurement	86
Nominal, Ordinal, Interval, and Ratio	87
More About the FREQUENCIES Procedure	88
A Frequency Table	89
A Bar Chart and a Frequency Table	89
A Bar Chart without a Frequency Table	89
A Frequency Table in a Condensed Form	89
What's Next?	89

8**SUMMARIZING DATA**

94

Descriptive Statistics	95
Nominal Variables	95
Ordinal Variables	96
Interval or Ratio Variables	97
Histograms	98
Differences between Bar Charts and Histograms	100
Uses of Histograms	100
More Descriptive Statistics	101
Other Percentiles	101
The Average or Arithmetic Mean	102
Mean, Median, or Mode?	102
How Much Do the Values Differ?	104
The Variance	105
The Standard Deviation	107
Computing the Variance with SPSS/PC+	107
More About the FREQUENCIES Procedure	108
The Simplest Way	108
No Frequency Table	108
A Histogram without a Frequency Table	108
Additional Statistics	108
Percentiles	109
What's Next?	109

9**COUNTING RESPONSES FOR COMBINATIONS OF VARIABLES**

114

Two Frequency Tables	115
Cross-Classification Tables	116
Percentages	117
Results of Percentaging	118
Column Percentages and Row Percentages	118

Dependent and Independent Variables	119
An Example: Is Marriage Exciting?	120
More Than Two Variables	121
Control Variables	123
More About the CROSSTABS Procedure	123
Making the Table	124
More than Two Variables in a Table	124
Calculating Percentages	124
Options for CROSSTABS	125
What's Next?	125

10 **CHANGING THE CODING SCHEME** 132

The RECODE Command	133
Changing Individual Values	133
Collapsing Values	135
How Does It Work?	136
Specifying the RECODE Command	136
Age and Excitement with Life	137
More About RECODE	139
Recoding Several Variables	139
What Happens to the Original Data?	139
What ELSE?	140
Missing Values	140
Lowest and Highest Values	141
Overlapping Values	141
Summary of Keywords Used in RECODE	142
What's Next?	142

11 **LOOKING AT MEANS** 144

Comparing Averages	145
Simple Solutions	148
More About the MEANS Procedure	148
Order of Variables	148
More Than One Table	148
Options for MEANS	149
What's Next?	149

12 **MODIFYING DATA VALUES** 154

An Example: Computing Age	155
The COMPUTE Command	155
How Does It Work?	156
Missing Values and COMPUTE	157
The IF Command	158
How Does It Work?	159
Missing Values and IF	160
When IFs Multiply	160
The Order of Transformation Commands	161
Labeling New Variables	161

Selecting and Sampling Cases	162
The SELECT IF Command	162
The SAMPLE Command	162
More About Transformations	163
String Variables	163
Temporary Case Selection	164
Functions	164
What's Next?	165

Part 3

13	MEANS FROM SAMPLES	168
	From Sample to Population	169
	Problems in Generalizing	170
	Sampling Variability	170
	A Computer Model	171
	Building 1,000 Samples	171
	Comparing the Results	171
	The Effect of Sample Size	174
	The Effect of Population Variability	175
	Other Statistics	176
	Distribution of Statistics	176
	What's Next?	177
14	WORKING WITH THE NORMAL DISTRIBUTION	180
	The Normal Distribution	181
	Areas in the Normal Distribution	182
	Standard Scores	182
	A Sample from the Normal Distribution	184
	Distributions that Aren't Normal	184
	More on the Distribution of the Means	186
	The Central Limit Theorem	186
	More About Means of Means	188
	The Standard Error of the Mean	189
	Calculating a Confidence Interval	190
	Smarter than Average?	193
	What's Next?	194
15	TESTING HYPOTHESES ABOUT TWO INDEPENDENT MEANS	198
	Is the Difference Real?	199
	Review of the Distribution of Means	200
	Differences between Means	200
	Evaluating a Difference between Means	202
	Drawing a Conclusion	204
	More on Hypothesis Testing	205
	Why Is That So Complicated?	205

T-Tests	207
Using the <i>t</i> -Distribution	208
Two Types of Errors	209
Output from the T-TEST Procedure	210
Interpreting a T-TEST	211
An Analogy: Coin Flips	212
Observed Significance Levels	213
Tails and Significance Tests	214
The Hypothesis-Testing Process	215
Assumptions Needed	216
More About the T-TEST Procedure	217
What's Next?	218

16

TESTING HYPOTHESES ABOUT TWO DEPENDENT MEANS 222

Paired Designs	223
Advantages of Comparing Pairs of Subjects	224
Some Possible Problems	225
Paired Data and the Computer	225
Analyzing Paired Designs	226
The Paired <i>t</i> -Test	227
The Confidence Interval for a Difference	228
Assumptions	229
Education and Excitement	230
Significance Versus Importance	232
More About Paired T-Tests	233
More than Two Variables	233
Options for T-Test	233
What's Next?	234

17

TESTING HYPOTHESES ABOUT INDEPENDENCE 238

Crosstabulation Again	239
The Null Hypothesis for a Crosstabulation	240
Expected Frequencies	241
Comparing Observed and Expected Frequencies	241
The Chi-Square Statistic	242
Using SPSS/PC+ To Calculate Chi-Square	243
Results of the Test	244
Relationships Between Variables	245
What Is a Relationship?	245
Testing Independence	246
Chi-Square in Larger Tables	246
Chi-Square Test of a Relationship	247
Expected Frequencies and Chi-Square	248
Degrees of Freedom	248
Interpreting the Chi-Square Test	249
Sample Size and the Chi-Square Statistic	250
More About Hypothesis Testing in CROSSTABS	252
What's Next?	252

18**COMPARING SEVERAL MEANS** 258

Describing the Groups	259
The Distribution of the Responses	260
Are the Means Really Different?	263
Analysis of Variance	263
Necessary Assumptions	263
Partitioning the Variability	264
Analysis of Variance in SPSS/PC+ ONEWAY	265
Computing Within-Groups Variability	266
Computing Between-Groups Variability	267
Calculating the <i>F</i> -Ratio	267
Multiple Comparison Procedures	268
Multiple Comparisons in ONEWAY	269
More Complicated Analysis of Variance Designs	270
Interactions	271
Analysis of Variance in SPSS/PC+ ANOVA	271
More About the ONEWAY Procedure	272
Obtaining Multiple Comparisons	273
Available Options	273
Available Statistics	273
More About the ANOVA Procedure	273
Estimation Method	274
Suppressing High-order Interactions	274
Displaying Cell Means	274
What's Next?	274

Part 4

19**MEASURING ASSOCIATION** 280

The Strength of a Relationship	281
Why Not Chi-Square?	281
Measures of Association	282
Measures of Association for Nominal Variables	282
Measures Based on Chi-Square	283
Proportional Reduction in Error	285
Measures of Association for Ordinal Variables	290
Concordant and Discordant Pairs	290
Measures Based on Concordant and Discordant Pairs	291
Measures Involving Interval Data	293
Testing Hypotheses	293
More About Statistics for CROSSTABS	293
What's Next?	294

20**PLOTTING DATA** 298

Plotting Your Variables	299
Interpreting a Plot	301
Cases with Similar Values	301
Using a Control Variable on a Plot	302
Interpreting a Control Plot	302
Using Reference Lines	304

Why Plot?	305
More About the PLOT Procedure	305
What's Next?	307

21 INTERPRETING CORRELATION COEFFICIENTS 312

Types of Relationships	313
Perfect Relationships	314
More about Linear Relationships	316
A Weaker Linear Relationship	316
Correlation	318
The Pearson Correlation Coefficient	318
Calculating the Correlation Coefficient	320
Testing Hypotheses about the Correlation Coefficient	321
Assumptions about the Data	322
Examining Many Coefficients	322
Missing Values	323
More About the CORRELATION Procedure	324
What's Next?	325

22 CALCULATING SIMPLE REGRESSION LINES 328

Choosing the Best Line	329
The Equation of a Line	330
Predicting Values from the Regression Line	332
Choosing the Dependent Variable	333
Determining How Well the Line Fits	333
Correlating Predicted and Observed Values	333
Explaining Variability	334
More About Regression with the PLOT Procedure	335
What's Next?	335

23 TESTING REGRESSION HYPOTHESES 340

The Population Regression Line	341
Additional Assumptions	342
Some Hypotheses of Interest	344
Testing Hypotheses with the REGRESSION Command	344
Regression Coefficients	345
Confidence Intervals for Regression Coefficients	347
Goodness of Fit of the Model	348
Another Test for a Linear Relationship	348
Other Regression Statistics	350
More About the REGRESSION Procedure	350
Multiple Regression	351
What's Next?	351

24 ANALYZING RESIDUALS 356

Residuals	357
Looking at Residuals	357
Reading the Casewise Results	358
Judging the Size of the Residuals	358
Looking for Outliers	359

Checking Assumptions with Residuals	360
Normality	360
Constant Variance	362
Linearity	364
Independence	368
A Final Comment on Assumptions	368
More About the REGRESSION Procedure	369
Histograms	370
Scatterplots	370
Splitting the Sample	370
What's Next?	371

25

LOOKING BEYOND	376
Multivariate Statistical Techniques	376
Multiple Linear Regression	377
Discriminant Analysis	379
Log-Linear Models	380
Factor Analysis	380
Cluster Analysis	381
Testing Hypotheses about Many Means	381
There's More	382

Appendix A	
STUDENTWARE ERROR MESSAGES	383
Making Errors	383
Correcting Errors	383
Memory Problems	384
Common Error Messages and Warnings	384

Appendix B	
ANSWERS TO SELECTED EXERCISES	393

Appendix C	
MORE ADVICE ON USING THE SYSTEM	401
Possible Problems	401
The CONFIG.SYS File	401
Setting the PATH	402
Studentware with DOS 2	402
Using a Hard Disk	403
Installation	404
Operating from a Hard Disk	405
Submitting Commands from REVIEW	405
The SET Command	405
The Initialization Profile	406
The Studentware Data Cache	407
Rearranging the Diskettes	408
The Studentware Program Files	409
The Diskette Table File	411

Appendix D	
USING THIS BOOK WITH SPSS/PC+	412
SPSS/PC+	412
REVIEW	413
Glossary	413
Syntax Help	414
SPSS/PC+ V2.0	414
Menus	414
Help	414

Appendix E	
NONPARAMETRIC PROCEDURES:	
THE NPAR TESTS COMMAND	415
The Mann-Whitney Test	416
The Sign Test	418
The Runs Test	419
The One-Sample Chi-Square Test	420
Expected Frequencies	421
More About the NPAR TESTS Command	421
One-Sample Tests	422
Two-Sample Tests	423
Optional Statistics	425
Missing Values	425
Sampling	425

BIBLIOGRAPHY	426
---------------------	-----

INDEX	427
--------------	-----

PART ONE

Getting Started

1	INTRODUCTION	2
2	DESIGNING A STUDY	6
3	DESIGNING FORMS FOR STUDIES	20
4	SETTING UP YOUR DATA FILE	32
5	RUNNING SPSS/PC+	50
6	DEFINING YOUR DATA	60

1

Introduction

Perhaps you're reading this book because you're aware of the importance of collecting, analyzing, and interpreting data. You already know that in education, business, medicine, and other disciplines, data are pivotal in decision making. Perhaps you're about to begin on a research project, either by choice or assignment, and you want to do it well. You want to design a good study, analyze the results properly, and prepare a cogent report that summarizes what you've found. You also know that using a computer for data analysis is essential. You don't want to waste your time poring through data with a hand calculator. If this is the case, you can proceed to the section "About this Book," which describes the organization of the book.

Or perhaps you're reading this book because you have to. It's an assignment. The statistics course is a requirement. The university thinks you should know something about statistics. You're not sure. You think a course in medieval astrology may be just as useful and probably more interesting. You may or may not enjoy using the computer, but it seems that you must use it. Forced unions of authors and readers can be difficult for both; perhaps it will help ease our relationship if we look at some of the reasons you'll profit from learning about data analysis.

WHAT ARE DATA?

In common usage "data" are any materials that serve as a basis for drawing conclusions. Drawing conclusions from data is an activity in which everyone engages—bankers, scholars, politicians, doctors, and corporate presidents. In theory, we base our foreign policy, methods of treating diseases, and corporate marketing strategies on "data."

There are many sources for data. We can conduct our own surveys or experiments, look at information from surveys other people have conducted, or examine data from all sorts of existing records—such as stock transactions or election tallies.

But acquiring data is not enough. We must determine what conclusions are justified based on the data. That's known as "data analysis." People and organizations deal with data in many different ways. Some