



The  
SPSS GUIDE to

---

# Data Analysis

---

for SPSS/PC+<sup>TM</sup>, 2nd Edition

MARIJA J. NORUŠIS

The  
SPSS GUIDE to

---

# Data Analysis

---

for SPSS/PC+<sup>™</sup>, 2nd Edition

**MARIJA J. NORUŠIS**

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

**SPSS Inc.**

For more information about SPSS/PC+ Studentware Plus™, 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  
Tel: (312) 329-2400  
Fax: (312) 329-3668

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.

General notice: Other product names mentioned herein are used for identification purposes only and may be trademarks of their respective companies.

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

The SPSS Guide to Data Analysis for SPSS/PC+™, 2nd Edition

Copyright © 1991 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 95

ISBN 0-13-178443-9

Library of Congress Catalog Card Number: 91-066147

# 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.

With SPSS/PC+ 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 Book**

The book is divided into four main parts: Getting Started, 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 major 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. Appendix B contains selected answers to the exercises. Appendix A explains how to correct some of the errors commonly made by users of Studentware. Appendix A explains how to format sum-

mary statistics for presentation. Appendix F presents nonparametric procedures not covered in the main text.

## 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. I also wish to thank my students and other users of previous versions of this book and software for helpful suggestions. Finally, I wish to thank Liz Adams for her editorial contributions.

Marija J. Norušis

## Contacting SPSS Inc.

If you would like to be on our mailing list, write to us at one of the addresses below. We will send you a copy of our newsletter and let you know about SPSS Inc. activities in your area.

**SPSS Inc.**  
444 North Michigan Ave.  
Chicago, IL 60611  
Tel: (312) 329-2400  
Fax: (312) 329-3668

**SPSS Federal Systems**  
Courthouse Place  
2000 North 14th St.  
Suite 320  
Arlington, VA 22201  
Tel: (703) 527-6777  
Fax: (703) 527-6866

**SPSS Latin America**  
444 North Michigan Ave.  
Chicago, IL 60611  
Tel: (312) 494-3226  
Fax: (312) 494-3227

**SPSS Benelux BV**  
P.O. Box 115  
4200 AC Gorinchem  
The Netherlands  
Tel: +31.1830.36711  
Fax: +31.1830.35839

**SPSS GmbH Software**  
Rosenheimer Strasse 30  
D-81669 Munich  
Germany  
Tel: +49.89.4890740  
Fax: +49.89.4483115

**SPSS Hellas SA**  
5 Ventiri Street  
115 28 Athens  
Greece  
Tel: +30.1.7251925  
Fax: +30.1.7291915

**SPSS UK Ltd.**  
SPSS House  
5 London Street  
Chertsey  
Surrey KT16 8AP  
United Kingdom  
Tel: +44.932.566262  
Fax: +44.932.567020

**SPSS France SARM**  
72-74 Avenue Edouard  
Vaillant  
92100 Boulogne  
France  
Tel: +33.1.4699.9670  
Fax: +33.1.4684.0180

**SPSS Hispanoportuguesa S.L.**  
Paseo Pintor Rosales, 26-4  
28008 Madrid  
Spain  
Tel: +34.1.547.3703  
Fax: +34.1.548.1346

**SPSS Scandinavia AB**  
Gamla Brokatan 36-38  
4th Floor  
111 20 Stockholm  
Sweden  
Tel: +46.8.102610  
Fax: +46.8.102550

**SPSS Italia srl**  
via Collegio di Spagna, 5/2  
40123 Bologna  
Italy  
Tel: +39.51.234574  
Fax: +39.51.238446

**SPSS Israel Ltd**  
39 Hagalim Blvd.  
Herzlia 46725  
Israel  
Tel: +972.9.598900  
Fax: +972.9.598903

**SPSS India Private Ltd.**  
Ashok Hotel, Suite 223  
50B Chanakyapuri  
New Delhi 110 021  
India  
Tel: +91.11.600121 x1029  
Fax: +91.11.688.8851

**SPSS Asia Pacific Pte. Ltd.**  
10 Anson Road, #34-07  
International Plaza  
Singapore 0207  
Singapore  
Tel: +65.221.2577  
Fax: +65.221.9920

**SPSS Japan Inc.**  
2-2-22 Jingu-mae  
Shibuya-ku, Tokyo  
150 Japan  
Tel: +81.3.5474.0341  
Fax: +81.3.5474.2678

**SPSS Australasia Pty. Ltd.**  
121 Walker Street  
North Sydney, NSW 2060  
Australia  
Tel: +61.2.954.5660  
Fax: +61.2.954.5616

# Contents

Part 1    Getting Started

---

## 1

### **RUNNING SPSS/PC+: A QUICK EXAMPLE    2**

- Starting SPSS/PC+    3
  - Creating a Data Directory    3
- Running an Analysis    5
  - Entering Variable Names    5
  - Running Commands    7
  - The More Prompt    7
  - Labeling the Variables    8
  - Listing the Data    9
- The Input    9
- The Output    10
- Running Another Command    10
- Saving and Getting an SPSS/PC+ System File    11
- Leaving SPSS/PC+    12
- What's Next?    12
- Exercises    13

## 2

### **SPSS/PC+ COMMANDS AND SESSIONS    14**

- Commands and Specifications    15
  - Subcommands and Keywords    15
- Rules for Constructing Commands    16
- The SPSS/PC+ Session    17
- What's Next?    19
- Exercises    19

## 3

### **A DATA ANALYSIS SESSION    22**

- A Sample Session    23
  - Starting SPSS/PC+    23
  - Identifying Your Data    23
  - Analyzing Your Data    24
  - Saving Your Results    25
  - Saving Your Commands    26
  - Printing a File    26
- Typing as an Alternative to Pasting    27
- Finding a Command in the Menus    28
- Running SPSS/PC+ with Menus off    29
  - Entering Commands in Examples and Exercises    29
- Summary of Menu Features    29
- What's Next?    30
- Exercises    31

## 4

### **USING SPSS/PC+ EFFICIENTLY 34**

- Editing Text in the Scratch Pad 35
- Using Function Keys and Mini-Menus 36
  - Function Key Commands 38
  - Additional Help 39
  - Shortcuts 39
- Modifying Data Formats 40
- What's Next? 41
- Exercises 42

## 5

### **DEFINING DATA 44**

- Cases, Variables, and Values 45
  - Missing Values 46
  - Variable Names 46
- Reading the Data into SPSS/PC+ 46
  - Spreadsheets and Databases: The Translate Command 47
  - Text Files: the Data List Command 48
- Describing the Data 50
  - Missing Values 50
  - Variable Labels 50
- Saving and Retrieving System Files 51
- What's Next? 52
- Exercises 52

## 6

### **DESIGNING A STUDY 54**

- Asking a Question 55
  - What Information Do You Need? 56
- Defining a Population 56
  - Sampling 57
- Using Surveys 59
  - Conducting Your Own Survey 60
  - Analyzing an Existing Survey 60
- Designing Experiments 61
  - Random Assignment 61
  - "Blind" Experiments 63
- What's Next? 65
- Exercises 66

## 7

### **DESIGNING FORMS FOR STUDIES 68**

- The General Social Survey 69
  - Is Life Exciting? 69
- Designing the Form 70
  - Arranging the Form 70
  - Coding the Data 71
  - Tips on Form Design 74
- Collecting the Data 75
- Reviewing and Preparing Forms for Data Entry 76
- What's Next? 76
- Exercises 77

## 8

### COUNTING RESPONSES FOR A SINGLE VARIABLE 82

- Counting Frequencies 83
  - Interpreting a Frequency Table 84
  - Checking the Data 84
- Percentages 85
  - Valid Percentages 85
  - Bar Charts 86
  - Cumulative Percentages 87
- Levels of Measurement 88
- More about the FREQUENCIES Procedure 90
  - A Frequency Table 90
  - A Bar Chart and a Frequency Table 91
  - A Bar Chart without a Frequency Table 91
  - A Frequency Table in a Condensed Format 91
- What's Next? 91
- Exercises 92

## 9

### SUMMARIZING DATA 96

- Summarizing Variables 97
  - Nominal Variables 97
  - Ordinal Variables 98
  - Interval or Ratio Variables 99
- Histograms 99
  - Differences between Bar Charts and Histograms 102
  - Uses of Histograms 103
- Descriptive Statistics 104
  - Other Percentiles 104
  - The Average or Arithmetic Mean 105
  - Mean, Median, or Mode? 106
- How Much Do the Values Differ? 107
  - The Variance 108
  - The Standard Deviation 110
  - Computing the Variance with SPSS/PC+ 110
- More about the FREQUENCIES Procedure 111
  - Basic Statistics 111
  - No Frequency Table 111
  - Histograms without Frequency Tables 111
  - Additional Statistics 112
  - Percentiles 112
- What's Next? 113
- Exercises 114

## 10

### COUNTING RESPONSES FOR COMBINATIONS OF VARIABLES 118

- Two Frequency Tables 119
- Crosstabulations 120
  - Percentages 121
  - Results of Percentaging 122
  - Column Percentages and Row Percentages 122

- Dependent and Independent Variables 123
- An Example: Is Marriage Exciting? 124
- More than Two Variables 125
  - Control Variables 126
- More about the Crosstabs Procedure 127
  - Crosstabulations of Two Variables 127
  - More than Two Variables in a Table 127
  - Percentages 128
  - Cell Contents 128
- What's Next? 129
- Exercises 130

## 11

### CHANGING THE CODING SCHEME 134

- The RECODE Command 135
  - Changing Individual Values 135
  - Collapsing Values 137
  - How Does It Work? 137
  - Specifying the RECODE Command 138
- Age and Excitement with Life 138
- More about RECODE 141
  - Recoding Several Variables 141
  - What Happens to the Original Data? 141
  - What Else? 142
  - Missing Values 142
  - Lowest and Highest Values 143
  - Overlapping Values 143
- Summary of Keywords Used in RECODE 144
- What's Next? 144
- Exercises 145

## 12

### LOOKING AT MEANS 148

- Comparing Averages 149
- Simple Solutions 152
- More about the MEANS Procedure 152
  - Order of Variables 152
  - More than One Table 152
  - Options for MEANS 153
- What's Next? 153
- Exercises 154

## 13

### ADDITIONAL WAYS OF DISPLAYING DATA 158

- The Boxplot 159
  - Extreme Values 161
  - Why Use Boxplots? 162
- Stem-and-leaf Plots 162
- More about the EXAMINE Procedure 164
  - Statistics Displayed by EXAMINE 164
  - Plots 165
  - Identifying Extreme Cases 165
- What's Next? 165
- Exercises 166

## 14

### **MODIFYING DATA VALUES 170**

- An Example: Computing Age 171
- The COMPUTE Command 172
  - How Does It Work? 172
  - Missing Values and COMPUTE 173
- The IF Command 174
  - How Does It Work? 175
  - Missing Values and IF 176
  - When IFs Multiply 177
- The Order of Transformation Commands 177
- Labeling New Variables 178
- Selecting and Sampling Cases 178
  - The SELECT IF Command 178
  - The SAMPLE Command 179
- More about Transformations 180
  - String Variables 180
  - Temporary Case Selection 180
  - Functions 181
- What's Next? 182
- Exercises 183

#### Part 3    Testing Hypotheses

---

## 15

### **MEANS FROM SAMPLES 186**

- from Sample to Population 187
  - Problems in Generalizing 188
  - Sampling Variability 188
- A Computer Model 189
  - Building 1,000 Samples 189
  - Comparing the Results 189
  - The Effect of Sample Size 192
- The Effect of Population Variability 193
  - Other Statistics 194
- Distributions of Statistics 194
- What's Next? 195
- Exercises 196

## 16

### **WORKING WITH THE NORMAL DISTRIBUTION 198**

- The Normal Distribution 199
  - Areas in the Normal Distribution 200
- Standard Scores 200
  - Tables of the Standard Normal Distribution 202
- A Sample from the Normal Distribution 203
  - Distributions that Aren't Normal 204
- More on the Distribution of the Means 205
  - The Central Limit Theorem 205
- More about Means of Means 208

The Standard Error of the Mean 208  
Calculating a Confidence Interval 209  
    Smarter than Average? 212  
What's Next? 213  
Exercises 214

## 17

### TESTING HYPOTHESES ABOUT TWO INDEPENDENT MEANS 218

Is the Difference Real? 219  
    Review of the Distribution of Means 220  
    Differences between Means 221  
    Evaluating a Difference between Means 222  
    Drawing a Conclusion 224  
More on Hypothesis Testing 225  
    Why is that so Complicated? 226  
    Steps for Hypothesis Testing 227  
T Tests 227  
    Using the T Distribution 228  
    Two Types of Errors 230  
    Output from the T-TEST Procedure 231  
Interpreting a T Test 232  
    An Analogy: Coin Flips 233  
    Observed Significance Levels 234  
    Tails and Significance Tests 235  
The Hypothesis-Testing Process 236  
    Assumptions Needed 237  
More about the T-TEST Procedure 238  
What's Next? 239  
Exercises 240

## 18

### TESTING HYPOTHESES ABOUT TWO DEPENDENT MEANS 244

Paired Designs 245  
    Advantages of Comparing Pairs of Subjects 246  
    Some Possible Problems 247  
    Paired Data and SPSS/PC+ 247  
Analyzing Paired Designs 248  
    The Paired T Test 249  
    The Confidence Interval for a Difference 250  
    Assumptions 251  
    Education and Excitement 255  
Significance versus Importance 257  
More about Paired T Tests 258  
    More than Two Variables 258  
What's Next? 259  
Exercises 260

## 19

### TESTING HYPOTHESES ABOUT INDEPENDENCE 264

- Crosstabulation Again 265
- The Null Hypothesis for a Crosstabulation 266
  - Expected Frequencies 267
  - Comparing Observed and Expected Frequencies 267
- The Chi-Square Statistic 268
- Using SPSS/PC+ to Calculate Chi-Square 269
  - Results of the Test 270
- Relationships between Variables 271
  - What is a Relationship? 271
  - Testing Independence 272
- Chi-Square in Larger Tables 272
  - Chi-Square Test of a Relationship 273
  - Expected Frequencies and Chi-Square 274
  - Degrees of Freedom 275
  - Interpreting the Chi-Square Test 276
- Sample Size and the Chi-Square Statistic 276
- More about Hypothesis Testing in CROSSTABS 278
- What's Next? 279
- Exercises 280

## 20

### COMPARING SEVERAL MEANS 286

- Describing the Groups 287
  - The Distribution of the Responses 288
  - Are the Means Really Different? 289
- Analysis of Variance 291
  - Necessary Assumptions 291
  - Partitioning the Variability 291
  - Analysis of Variance in SPSS/PC+ ONEWAY 293
  - Computing Within-groups Variability 294
  - Computing Between-groups Variability 295
  - Calculating the  $F$  Ratio 295
- Multiple Comparison Procedures 296
  - Multiple Comparisons in ONEWAY 297
- More Complicated Analysis of Variance Designs 298
  - Interactions 299
  - Analysis of Variance in SPSS/PC+ ANOVA 299
- More about the ONEWAY Procedure 301
  - Obtaining Multiple Comparisons 301
  - Available Options 302
  - Available Statistics 302
- More about the ANOVA Procedure 302
  - Estimation Method 302
  - Suppressing High-order Interactions 303
  - Displaying Cell Means 303
- What's Next? 303
- Exercises 304

## **21 MEASURING ASSOCIATION 310**

- The Strength of a Relationship 311
  - Why Not Chi-Square? 311
  - Measures of Association 312
- Measures of Association for Nominal Variables 312
  - Measures Based on Chi-Square 313
  - Proportional Reduction in Error 315
- Measures of Association for Ordinal Variables 320
  - Concordant and Discordant Pairs 321
  - Measures Based on Concordant and Discordant Pairs 322
- Measures Involving Interval Data 324
- Testing Hypotheses 324
- More about Statistics for CROSSTABS 324
- What's Next? 326
- Exercises 327

## **22 PLOTTING DATA 331**

- Plotting Your Variables 331
  - Interpreting a Plot 333
  - Cases with Similar Values 333
- Using a Control Variable on a Plot 334
  - Interpreting a Control Plot 334
- Using Reference Lines 336
- Why Plot? 337
- More about the Plot Procedure 337
  - Multiple Plots 338
  - Reference Lines 338
  - Control Plots 338
  - Overlay Plots 339
  - Regression Plots 339
- What's Next? 339
- Exercises 340

## **23 INTERPRETING CORRELATION COEFFICIENTS 344**

- Types of Relationships 345
  - Perfect Relationships 346
  - More about Linear Relationships 347
  - Another Linear Relationship 348
- Correlation 350
  - The Pearson Correlation Coefficient 350
- Calculating the Correlation Coefficient 351
  - Testing Hypotheses about the Correlation Coefficient 353
  - Assumptions about the Data 354
  - Examining Many Coefficients 354
  - Missing Values 355
- More about the CORRELATION Procedure 356
  - Options and Statistics 356
- What's Next? 357
- Exercises 358

## 24

### **CALCULATING SIMPLE REGRESSION LINES 362**

- Choosing the Best Line 363
  - The Equation of a Line 364
  - Predicting Values from the Regression Line 365
  - Choosing the Dependent Variable 367
- Determining How Well the Line Fits 367
  - Predicted and Observed Values 368
  - Explaining Variability 369
- More about Regression with the PLOT Procedure 369
- What's Next? 370
- Exercises 371

## 25

### **TESTING REGRESSION HYPOTHESES 374**

- The Population Regression Line 375
  - Additional Assumptions 376
- Some Hypotheses of Interest 378
  - Testing Hypotheses with the REGRESSION Command 378
  - Regression Coefficients 378
  - Are the Population Values Zero? 380
  - Confidence Intervals for Regression Coefficients 380
- Goodness of Fit of the Model 381
  - Another Test for a Linear Relationship 383
  - Other Regression Statistics 383
- More about the REGRESSION Procedure 384
  - Multiple Regression 384
- What's Next? 385
- Exercises 386

## 26

### **ANALYZING RESIDUALS 390**

- Residuals 391
  - Looking at Residuals 391
  - Reading the Casewise Results 392
  - Judging the Size of the Residuals 392
  - Looking for Outliers 393
- Checking Assumptions with Residuals 394
  - Normality 394
  - Constant Variance 396
  - Linearity 398
  - Independence 402
  - A Final Comment on Assumptions 402
- More about the Regression Procedure 403
  - Casewise Plots and Temporary Variables 403
  - Histograms 404
  - Scatterplots 404
  - Splitting the Sample 404
- What's Next? 405
- Exercises 405

**LOOKING BEYOND 410**

- Multivariate Statistical Techniques 411
  - Multiple Linear Regression 412
  - Discriminant Analysis 414
  - Log-linear Models 414
  - Factor Analysis 415
  - Cluster Analysis 416
  - Testing Hypotheses about Many Means 416
- There's More 417

Part 5    **Appendixes**
**Appendix A**  
**ERROR MESSAGES 421**
**Appendix B**  
**ANSWERS TO SELECTED EXERCISES 440**
**Appendix C**  
**MORE ADVICE ON USING THE SYSTEM 450**

- Possible Problems 450
  - The config.sys File 450
  - Setting the Path 451
- The SET Command 452
  - The Initialization Profile 453
- Editing a Text File in SPSS/PC+ 453
  - Creating a New File 454
  - Loading an Existing Text File 454
  - Editing the File 454
  - Saving the File (Whole or Part) 455
- Printing Results While Running 455
- Running SPSS/PC+ from the Prompt 455
- Finding Commands in SPSS/PC+ 456

**Appendix D**  
**FORMATTING INFORMATION WITH REPORT 457**

- A Basic Report 457
- More about the REPORT Command 465
  - The FORMAT Subcommand 466
  - The OUTFILE Subcommand 466
  - The VARIABLES Subcommand 466
  - The BREAK Subcommand 467
  - The SUMMARY Subcommand 467
  - Titles and Footnotes 469
  - Missing Values 469

## **Appendix E**

### **NONPARAMETRIC PROCEDURES: THE NPAR TESTS COMMAND 471**

The Mann-Whitney Test 472

The Sign Test 474

The Runs Test 475

The One-Sample Chi-Square Test 476

Expected Frequencies 477

More about the NPAR TESTS Command 477

One-Sample Tests 478

Two-Sample Tests 480

Optional Statistics 481

Missing Values 482

Subsampling 482

## **Appendix F**

### **AREAS UNDER THE NORMAL CURVE 483**

## **Appendix G**

### **THE $t$ DISTRIBUTION 484**

## **BIBLIOGRAPHY 486**

## **INDEX 489**

## PART ONE

# Getting Started

- 1** RUNNING SPSS/PC+: A QUICK EXAMPLE 2
- 2** SPSS/PC+ COMMANDS AND SESSIONS 14
- 3** A DATA ANALYSIS SESSION 22
- 4** USING SPSS/PC+ EFFICIENTLY 34
- 5** DEFINING DATA 44
- 6** DESIGNING A STUDY 54
- 7** DESIGNING FORMS FOR STUDIES 68