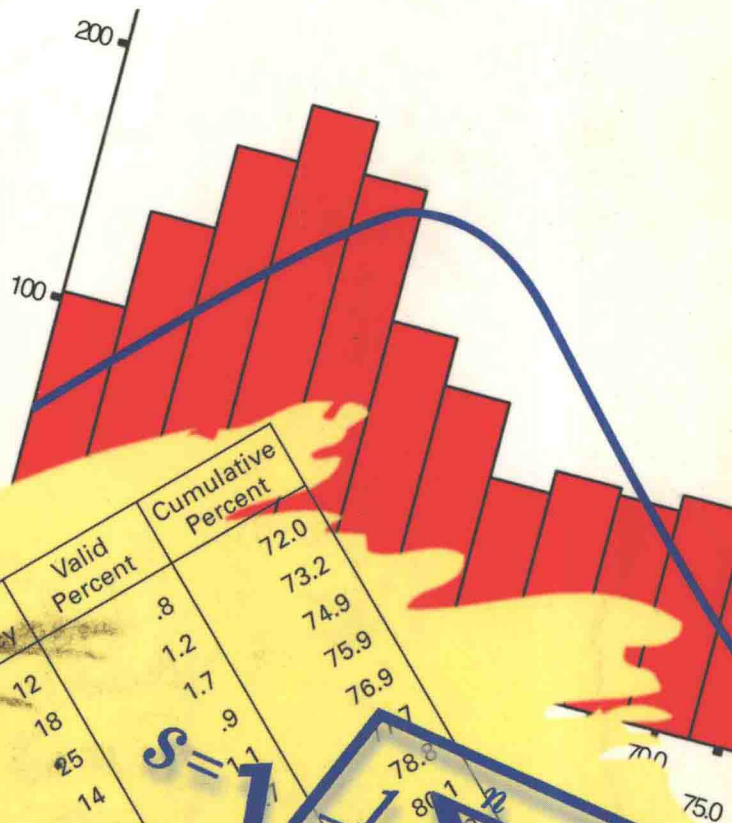


SPSS® 8.0

Guide to Data Analysis

Marija J. Norušis

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$



	Valid Percent	Cumulative Percent
5	.3	.3
17	1.1	1.5
18	1.2	2.7
22	1.5	4.1
15	1.0	5.2
28	1.9	7.0
23	1.5	8.6
	2.0	10.6
	1.8	12.4
		13.8
		16.7

	Valid	Frequency	Valid Percent	Cumulative Percent
56	12	.8		
57	18	1.2		
58	25	1.7		
59	14	.9		
60	16	1.1		
61	11	.7		
62	17	1.2		
63	19	.9		
64	13	1.1		
65	17	1.3		
66	19	.7		
67	11	1.1		
68	16	1.1		
69	10			
	9			
	15			

$$s = 1.1$$

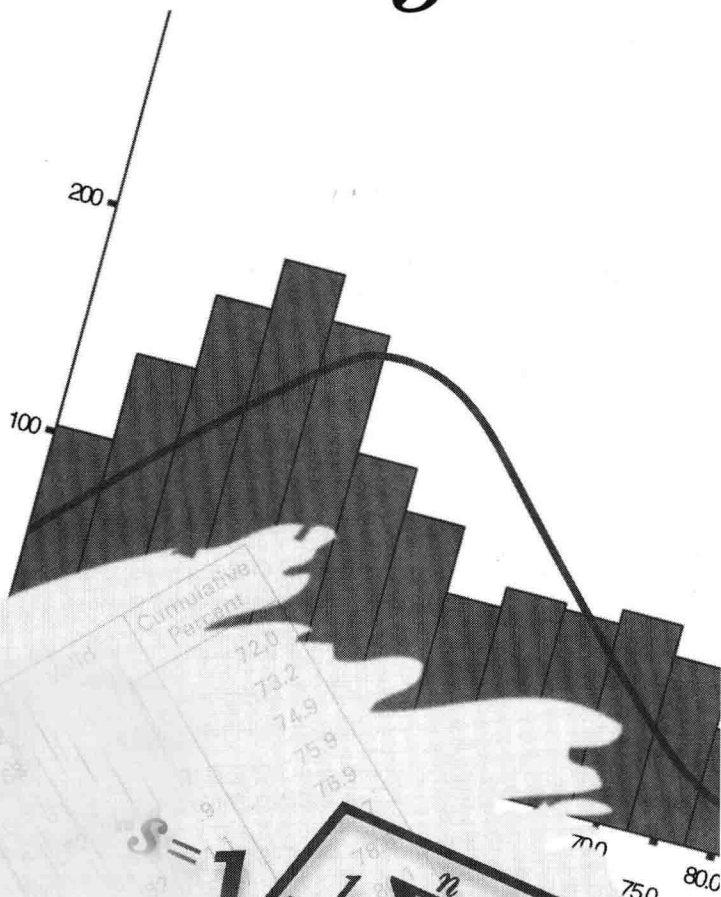
SPSS®

SPSS® 8.0

Guide to Data Analysis

Marija J. Norušis

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$



$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

PRENTICE HALL
Upper Saddle River, New Jersey 07458

SPSS®

For more information about SPSS® software products, please visit our WWW site at <http://www.spss.com> or contact:

Marketing Department
SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6307
Tel: (312) 651-3000
Fax: (312) 651-3668

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., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6307

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

TableLook is a trademark of SPSS Inc.

ImageStream® Graphics & Presentation Filters, copyright © 1991-1997 by INSO Corporation. All Rights Reserved.

ImageStream Graphics Filters is a registered trademark and ImageStream is a trademark of INSO Corporation.

Windows is a registered trademark of Microsoft Corporation.

SPSS® 8.0 Guide to Data Analysis
Published by Prentice-Hall, Inc., 1998
Upper Saddle River, New Jersey 07458

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.

ISBN 0-13-687484-3

Preface

The *SPSS Guide to Data Analysis* has weathered many transitions as SPSS software has evolved from mainframes to PC's to Windows. The current book and software bear little resemblance to their shaggy mainframe ancestors (although I like to think that the book was always friendlier than the software!). What has not changed is the goal of the book—to provide an unintimidating introduction both to data analysis and to SPSS. This book can be used either as a supplementary text or as a primary text in an introductory course in data analysis. It is designed for use with SPSS 8.0, including the Student Version, running on Windows 95 or Windows NT 4.0.

Data Files

Since the best way to learn about data analysis is to actually do it, this book uses real data to solve a variety of problems. Data from the General Social Survey are analyzed in several of the chapters and in the exercises. Examples and exercises are not limited to the General Social Survey, however. The variety of data files used should make this book appealing to a broader range of students and teachers. The data files used in the chapters and in the exercises are supplied on the disk included with this book. They can also be downloaded free of charge from the SPSS World Wide Web site at <http://www.spss.com> or via anonymous FTP at <ftp.spss.com>. The file location on the FTP site is `\pub\spss\sample\datasets` and the filename is *norusis.exe*.

Using This Book

This book is divided into four parts: “Getting Started with SPSS,” “Describing Data,” “Testing Hypotheses,” and “Examining Relationships.” Within these parts, each chapter presents a problem and statistical techniques useful for solving it. Detailed descriptions of the SPSS procedures used are also provided. Each chapter also contains extensive exercises that use the data files described above. Instructors can choose exercises that best suit the interests and sophistication of their students.

Send Me Comments

I am always eager to receive feedback and suggestions from readers of my books. Please send your suggestions and comments about the book (not the software) to *MJNORUSIS@AOL.COM*.

Acknowledgments

I wish to thank the many users of the *Guide to Data Analysis*, in its various incarnations and editions, for their helpful comments and suggestions. In particular, I wish to thank my students, who always made sure I knew when something I thought was admirably clear was not. I also wish to thank the reviewers of this book, especially Susan Shott, who made many valuable recommendations. Members of the Publications department at SPSS Inc. were involved in many aspects of the production of this book. I am grateful for their patience and skill. Finally, I wish to thank my husband, Bruce Stephenson, for continuing our multidimensional collaboration, and Daina Norušis for providing a constant stream of excellent reasons not to waste my time on boring books.

Marija J. Norušis

Contents

Part 1 **Getting Started with SPSS**

1 Introduction 1

- About This Book 2
 - Getting Started with SPSS 2
 - Describing Data 2
 - Testing Hypotheses 3
 - Examining Relationships 3
 - Let's Get Started 4

2 An Introductory Tour: SPSS for Windows 5

- Starting SPSS for Windows 5
 - Help Is Always at Hand 7
- Copying the Data Diskette 7
- Opening a Data File 8
- Statistical Procedures 11
 - The Viewer Window 14
 - Viewer Objects 16
- The Data Editor Window 21
 - Entering Non-Numeric Data 23
 - Clearing the Data Editor without Saving Changes 24
- The SPSS Online Tutorial 25
- The SPSS Toolbar 26
- The SPSS Help System 27
 - Contextual Help 29
- What's Next? 29

Part 2 Describing Data

3 Counting Responses 31

Describing Variables 32

A Simple Frequency Table 33

Sorting Frequency Tables 38

Pie Charts 39

Bar Charts 40

Summarizing the Age Variable 41

Histograms 43

Mode and Median 44

Percentiles 46

Summary 47

What's Next? 47

How to Obtain a Frequency Table 48

Format: Appearance of the Frequency Table 49

Statistics: Univariate Statistics 49

Charts: Bar Charts, Pie Charts, and Histograms 50

Exercises 51

4 Computing Descriptive Statistics 59

Summarizing Data 60

Scales of Measurement 60

Mode, Median, and Arithmetic Average 62

Comparing Mean and Median 63

Measures of Variability 64

Range 64

Variance and Standard Deviation 65

The Coefficient of Variation 66

Standard Scores 67

Summary 69

What's Next? 69

How to Obtain Univariate Descriptive Statistics 69

Options: Choosing Statistics and Sorting Variables 71

Exercises 72

5 Comparing Groups 77

Education and Job Satisfaction 78

Plotting Mean and Standard Deviation 78

Layers: Defining Subgroups by More than One Variable 79

Summary 82

What's Next? 82

How to Obtain Subgroup Means 83

Layers: Defining Subgroups by More than One Variable 84

Options: Additional Statistics and Display of Labels 85

Exercises 86

6 Looking at Distributions 91

Age and Job Satisfaction 91

Identifying Extreme Values 93

Percentiles 95

Plots 96

Histograms and Stem-and-Leaf Plots 96

Boxplots 100

Summary 102

What's Next? 102

How to Explore Distributions 102

Explore Statistics 104

Graphical Displays 104

Options 106

Exercises 107

7 Counting Responses for Combinations of Variables 111

Income and Job Satisfaction 112

Row and Column Percentages 114

Bar Charts 117

Adding Control Variables 119

Summary 121

What's Next? 121

How to Obtain a Crosstabulation 122

Layers: Three or More Variables at Once 123

Cells: Percentages, Expected Counts, and Residuals 124

Bivariate Statistics 125

Format: Adjusting the Table Format 125

Exercises 126

8 Plotting Data 133

Examining Population Indicators 134

Simple Scatterplots 134

Sunflower Plots 138

Scatterplot Matrices 139

Overlay Plots 140

Three-Dimensional Plots 143

Identifying Unusual Points 146

Rotating 3-D Scatterplots 147

Summary 148

What's Next? 148

How to Obtain a Scatterplot 148

Obtaining a Simple Scatterplot 149

Obtaining an Overlay Scatterplot 150

Obtaining a Scatterplot Matrix 151

Obtaining a 3-D Scatterplot 152

Scatterplot Options 153

Case Identification 156
Rotating a 3-D Scatterplot 158
Exercises 159

Part 3 Testing Hypotheses

9 Evaluating Results from Samples 163

From Sample to Population 164
 A Computer Model 164
 The Effect of Sample Size 169
 The Binomial Test 171

Summary 174
What's Next? 174
 Exercises 175

10 The Normal Distribution 177

The Normal Distribution 177
 Samples from a Normal Distribution 181
 Means from a Normal Population 182
 Are the Sample Results Unlikely? 184
 Testing a Hypothesis 186
 Means from Non-Normal Distributions 186
 Means from a Uniform Distribution 187

Summary 189
What's Next? 189
 Exercises 190

11 Testing a Hypothesis about a Single Mean 197

- The Mythical Work Week 198
 - Examining the Data 198
- The T Distribution 200
 - Calculating the T Statistic 202
- Confidence Intervals 203
 - Other Confidence Levels 207
 - Confidence Interval for a Difference 207
 - Confidence Intervals and Hypothesis Tests 208
- Null Hypotheses and Alternative Hypotheses 208
 - Rejecting the Null Hypothesis 210
- Summary 211
- What's Next? 211
- How to Obtain a One-Sample T Test 212
 - Options: Confidence Level and Missing Data 212
 - Exercises 213

12 Testing a Hypothesis about Two Related Means 217

- Marathon Runners in Paired Designs 218
 - Looking at Differences 219
 - Is the Mean Difference Zero? 221
 - Two Approaches 221
- The Paired-Samples T Test 223
 - Are You Positive? 224
 - Some Possible Problems 225
 - Examining Normality 225
- Summary 227
- What's Next? 228
- How to Obtain a Paired-Samples T Test 228
 - Options: Confidence Level and Missing Data 229
 - Exercises 230

13	Testing a Hypothesis about Two Independent Means	233
	Looking at Age Differences	234
	Descriptive Statistics	235
	Distribution of Differences	235
	Standard Error of the Mean Difference	236
	Computing the T Statistic	236
	Output from the Two Independent-Samples T Test	237
	Confidence Intervals for the Mean Difference	238
	Testing the Equality of Variances	240
	Comparing Education	241
	Can You Prove the Null Hypothesis?	243
	Interpreting the Observed Significance Level	244
	Power	245
	Monitoring Death Rates	246
	Does Significant Mean Important?	249
	Summary	249
	What's Next?	249
	How to Obtain an Independent-Samples T Test	250
	Define Groups: Specifying the Subgroups	251
	Options: Confidence Level and Missing Data	252
	Exercises	253

14	One-Way Analysis of Variance	259
	Hours in a Work Week	260
	Describing the Data	260
	Confidence Intervals for the Group Means	261
	Testing the Null Hypothesis	262
	Assumptions Needed for Analysis of Variance	263
	Analyzing the Variability	264
	Comparing the Two Estimates of Variability	266
	The Analysis-of-Variance Table	267

Multiple Comparison Procedures 270
Summary 274
What's Next? 274
How to Obtain a One-Way Analysis of Variance 274
 Post Hoc Multiple Comparisons: Finding the Difference 276
 Options: Statistics and Missing Data 276
 Exercises 277

15 Two-Way Analysis of Variance 281

The Design 282
 Examining the Data 283
 Testing Hypotheses 285
 Degree and Gender Interaction 289
 Necessary Assumptions 290
 Analysis-of-Variance Table 291
 Testing the Degree-by-Gender Interaction 292
 Testing the Main Effects 293
 Removing the Interaction Effect 294
 Where Are the Differences? 296
Multiple Comparison Results 296
 Checking Assumptions 297
Extensions 299
Summary 300
What's Next? 300
How to Obtain a Factorial Analysis of Variance 301
 GLM General Factorial: Model 302
 GLM General Factorial: Plots 303
 GLM General Factorial: Post Hoc 304
 GLM General Factorial: Options 304
 GLM General Factorial: Save 305
 Exercises 306

16 Comparing Observed and Expected Counts 309

- Education and Anomia 309
 - Observed and Expected Counts 310
 - The Chi-Square Statistic 312
- College Degrees and Perception of Life 315
 - A Larger Table 316
- A One-Sample Chi-Square Test 318
- Power Concerns 320
- Summary 321
- What's Next? 321
 - Exercises 322

17 Nonparametric Tests 325

- Nonparametric Tests for Paired Data 326
 - Sign Test 326
 - McNemar's Test 328
 - Wilcoxon Test 330
- Mann-Whitney Test 332
- Kruskal-Wallis Test 334
- Runs Test 335
- Summary 337
- How to Obtain Nonparametric Tests 337
 - Chi-Square Test 337
 - Binomial Test 339
 - Runs Test 340
 - Two-Independent-Samples Tests 341
 - Several-Independent-Samples Tests 342
 - Two-Related-Samples Tests 343
 - Options: Descriptive Statistics and Missing Values 345
 - Exercises 346

Part 4 **Examining Relationships**

18 Measuring Association 349

- The Strength of a Relationship 350
 - Measures of Association 351
- Measures of Association for Nominal Variables 352
 - Measures Based on Chi-Square 352
 - Proportional Reduction in Error 354
- Measures of Association for Ordinal Variables 360
 - Concordant and Discordant Pairs 361
 - Measures Based on Concordant and Discordant Pairs 362
 - Correlation-Based Measures 365
 - Measuring Agreement 366
- Summary 368
- What's Next? 368
 - Exercises 369

19 Linear Regression and Correlation 373

- Life Expectancy and Birthrate 374
 - Choosing the “Best” Line 375
- Calculating the Least-Squares Line 380
 - Calculating Predicted Values and Residuals 382
 - Determining How Well the Line Fits 383
 - Explaining Variability 387
 - Some Warnings 389
- Summary 391
- What's Next? 391
- How to Obtain a Linear Regression 392
 - Statistics: Further Information on the Model 393
 - Residual Plots: Basic Residual Analysis 394

Linear Regression Save: Creating New Variables 396
Linear Regression Options 398
Exercises 399

20 Testing Regression Hypotheses 407

The Population Regression Line 407
 Assumptions Needed for Testing Hypotheses 408
Testing Hypotheses 409
 Testing that the Slope Is Zero 410
 Confidence Intervals for the Slope and Intercept 412
Predicting Life Expectancy 412
 Predicting Means and Individual Observations 413
 Standard Error of the Predicted Mean 414
 Confidence Intervals for the Predicted Means 415
 Prediction Intervals for Individual Cases 417
Summary 419
What's Next? 419
How to Obtain a Bivariate Correlation 419
 Options: Additional Statistics and Missing Data 421
How to Obtain a Partial Correlation 422
 Options: Additional Statistics and Missing Data 423
Exercises 424

21 Analyzing Residuals 431

Residuals 432
 Standardized Residuals 433
 Studentized Residuals 434
 Checking for Normality 435
 Checking for Constant Variance 438
 Checking Linearity 440
Checking Independence 443

A Final Comment on Assumptions	444
Looking for Influential Points	445
Studentized Deleted Residuals	447
Summary	448
What's Next?	448
Exercises	449

22 Building Multiple Regression Models 455

Predicting Life Expectancy	456
The Model	456
Assumptions for Multiple Regression	457
Examining the Variables	458
Looking at How Well the Model Fits	460
Examining the Coefficients	462
Interpreting the Partial Regression Coefficients	464
Changing the Model	465
Partial Correlation Coefficients	466
Tolerance and Multicollinearity	467
Beta Coefficients	468
Building a Regression Model	469
Methods for Selecting Variables	470
Summary	478
What's Next?	478
How to Obtain a Multiple Linear Regression	479
Options: Variable Selection Criteria	480
Exercises	482

23 Multiple Regression Diagnostics 489

Examining Normality	490
Scatterplots of Residuals	492