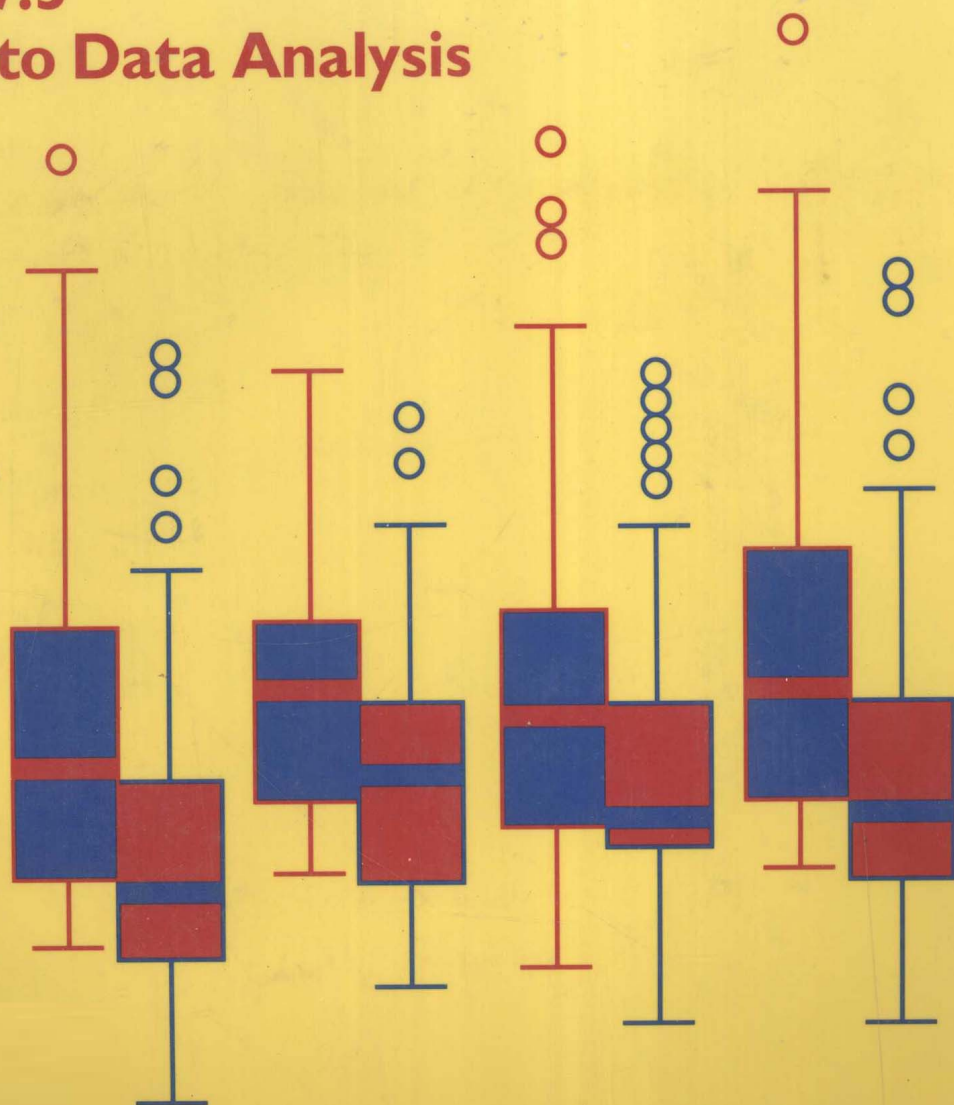


USED

# SPSS

SPSS® 7.5

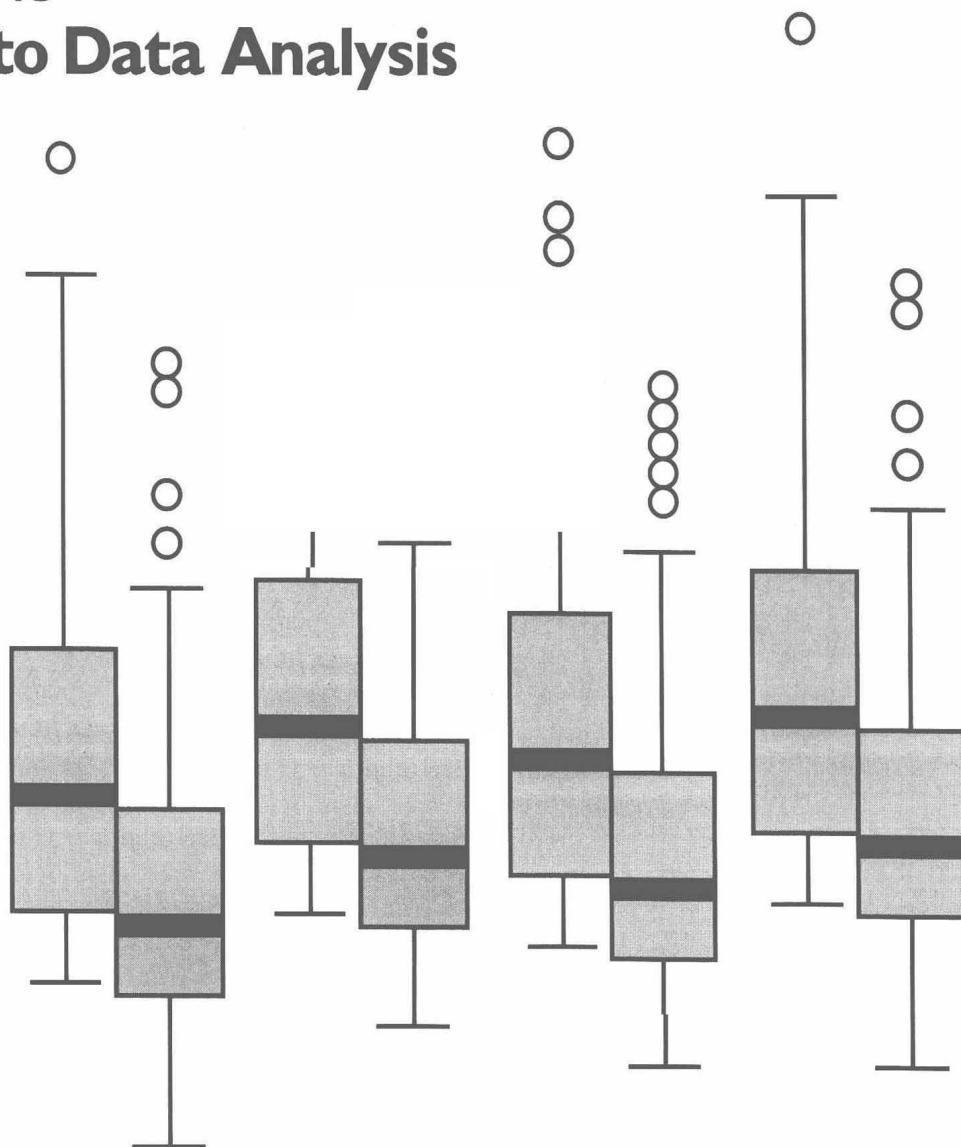
Guide to Data Analysis



Marija J. Norušis

# SPSS

## SPSS® 7.5 Guide to Data Analysis



**Marija J. Norušis**



PRENTICE HALL  
Upper Saddle River, New Jersey 07458

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

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.

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® 7.5 Guide to Data Analysis



Published by Prentice-Hall, Inc., 1997

A Simon & Schuster Company

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.

3 4 5 6 7 8 9 0 02 01 00 99 98

ISBN 0-13-656877-7

Library of Congress Catalog Card Number: 97-066250

# Preface

Although this book shares the same title as a series of books for other platforms and operating environments, it is fully up to date for SPSS 7.5. I hope that the many users of the earlier books will find it not just changed but improved. Its goal remains the same: to provide an unimintimidating introduction both to data analysis and to SPSS. This book can be used either as a supplementary text or as the primary text in an introductory course in data analysis. It is designed for use with SPSS 7.5, including the student version, in the Windows 95 operating environment.

This product, built around a graphical user interface and featuring an innovative Output Navigator, represents a substantial change from earlier versions of SPSS. Students no longer have to learn arcane commands. Classroom time no longer has to be spent on teaching students how to make things work and what to do when they don't work. There is no such thing as a job that won't run! In addition, the graphical capabilities of the software have been greatly expanded, and the statistical output of the program can be handled with as much flexibility as the graphics. As a natural result, both teaching and learning about data analysis are easier and a lot more interesting than they once were.

## 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 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.

## 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. Jonathan Healy and Bonnie Melton, of SPSS Inc., were involved in many aspects of the preparation 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
- Opening a Data File 7
- Statistical Procedures 9
  - The Output Navigator 12
  - Navigator Objects 14
- The Data Editor Window 19
  - Entering Non-Numeric Data 20
  - Clearing the Data Editor without Saving Changes 22
- The SPSS Online Tutorial 23
- The SPSS Toolbar 24
- The SPSS Help System 24
  - Contextual Help 26
- What's Next? 27

## **Part 2 Describing Data**

### **3 Counting Responses 29**

Describing Variables 30

A Simple Frequency Table 31

Sorting Frequency Tables 35

Pie Charts 36

Bar Charts 37

Summarizing the Age Variable 38

Histograms 40

Mode and Median 41

Percentiles 43

Summary 44

What's Next? 44

How to Obtain a Frequency Table 45

Format: Appearance of the Frequency Table 46

Statistics: Univariate Statistics 46

Charts: Bar Charts, Pie Charts, and Histograms 47

Exercises 48

### **4 Computing Descriptive Statistics 57**

Summarizing Data 58

Scales of Measurement 58

Mode, Median, and Arithmetic Average 60

Comparing Mean and Median 61

Measures of Variability 62

Range 62

Variance and Standard Deviation 63

The Coefficient of Variation 64

Standard Scores 65

Summary	67
What's Next?	67
How to Obtain Univariate Descriptive Statistics	67
Options: Choosing Statistics and Sorting Variables	69
Exercises	70

## **5 Comparing Groups 75**

Education and Job Satisfaction	76
Plotting Mean and Standard Deviation	76
Layers: Defining Subgroups by More than One Variable	77
Summary	80
What's Next?	80
How to Obtain Subgroup Means	81
Layers: Defining Subgroups by More than One Variable	82
Options: Additional Statistics and Display of Labels	83
Exercises	84

## **6 Looking at Distributions 89**

Age and Job Satisfaction	89
Identifying Extreme Values	91
Percentiles	93
Plots	94
Histograms and Stem-and-Leaf Plots	94
Boxplots	98
Summary	100
What's Next?	100
How to Explore Distributions	100
Explore Statistics	102
Graphical Displays	102
Options	104
Exercises	105



## **7 Counting Responses for Combinations of Variables 109**

Income and Job Satisfaction 110

Row and Column Percentages 112

Bar Charts 115

Adding Control Variables 117

Summary 119

What's Next? 119

How to Obtain a Crosstabulation 120

Layers: Three or More Variables at Once 121

Cells: Percentages, Expected Counts, and Residuals 122

Bivariate Statistics 123

Format: Adjusting the Table Format 123

Exercises 124

## **8 Plotting Data 131**

Examining Population Indicators 132

Simple Scatterplots 132

Sunflower Plots 136

Scatterplot Matrices 137

Overlay Plots 138

Three-Dimensional Plots 141

Identifying Unusual Points 144

Rotating 3-D Scatterplots 145

Summary 146

What's Next? 146

How to Obtain a Scatterplot 146

Obtaining a Simple Scatterplot 147

Obtaining an Overlay Scatterplot 148

Obtaining a Scatterplot Matrix 149

Obtaining a 3-D Scatterplot 150

Scatterplot Options 151  
Case Identification 154  
Rotating a 3-D Scatterplot 156  
Exercises 157

## **Part 3 Testing Hypotheses**

### **9 Evaluating Results from Samples 161**

From Sample to Population 162  
    A Computer Model 162  
    The Effect of Sample Size 167  
    The Binomial Test 169  
Summary 172  
What's Next? 172  
    Exercises 173

### **10 The Normal Distribution 175**

The Normal Distribution 175  
    Samples from a Normal Distribution 179  
    Means from a Normal Population 180  
    Are the Sample Results Unlikely? 182  
    Testing a Hypothesis 184  
    Means from Non-Normal Distributions 184  
    Means from a Uniform Distribution 185  
Summary 187  
What's Next? 187  
    Exercises 188

## **11 Testing a Hypothesis about a Single Mean 195**

The Mythical Work Week 196

Examining the Data 196

The T Distribution 198

Calculating the T Statistic 200

Confidence Intervals 201

Other Confidence Levels 205

Confidence Interval for a Difference 205

Confidence Intervals and Hypothesis Tests 206

Null Hypotheses and Alternative Hypotheses 206

Rejecting the Null Hypothesis 208

Summary 209

What's Next? 209

How to Obtain a One-Sample T Test 210

Options: Confidence Level and Missing Data 210

Exercises 211

## **12 Testing a Hypothesis about Two Related Means 215**

Marathon Runners in Paired Designs 216

Looking at Differences 217

Is the Mean Difference Zero? 219

Two Approaches 219

The Paired-Samples T Test 221

Are You Positive? 222

Some Possible Problems 223

Examining Normality 223

Summary 225

What's Next? 226

How to Obtain a Paired-Samples T Test 226

Options: Confidence Level and Missing Data 227

Exercises 228

## **13 Testing a Hypothesis about Two Independent Means 231**

- Looking at Age Differences 232
  - Descriptive Statistics 233
  - Distribution of Differences 233
  - Standard Error of the Mean Difference 234
  - Computing the T Statistic 234
  - Output from the Two Independent-Samples T Test 235
  - Confidence Intervals for the Mean Difference 236
  - Testing the Equality of Variances 238
- Comparing Education 239
  - Can You Prove the Null Hypothesis? 241
  - Interpreting the Observed Significance Level 242
  - Power 243
  - Monitoring Death Rates 244
  - Does Significant Mean Important? 247
- Summary 247
- What's Next? 247
- How to Obtain an Independent-Samples T Test 248
  - Define Groups: Specifying the Subgroups 249
  - Options: Confidence Level and Missing Data 250
  - Exercises 251

## **14 One-Way Analysis of Variance 257**

- Hours in a Work Week 258
  - Describing the Data 258
  - Confidence Intervals for the Group Means 259
  - Testing the Null Hypothesis 260
  - Assumptions Needed for Analysis of Variance 261
  - Analyzing the Variability 262
  - Comparing the Two Estimates of Variability 264
  - The Analysis-of-Variance Table 265

Multiple Comparison Procedures	268
Summary	272
What's Next?	272
How to Obtain a One-Way Analysis of Variance	272
Post Hoc Multiple Comparisons: Finding the Difference	274
Options: Statistics and Missing Data	274
Exercises	275

## **15 Two-Way Analysis of Variance 279**

The Design	280
Examining the Data	281
Testing Hypotheses	282
Necessary Assumptions	286
Analysis-of-Variance Table	286
Testing the Degree-by-Gender Interaction	288
Testing the Main Effects	289
Putting It All Together	289
Extensions	290
Summary	291
What's Next?	291
How to Obtain a Two-Way Analysis of Variance	292
Define Range: Specifying the Subgroups	293
Options: Analysis Method, Statistics, and Interactions	294
Exercises	295

## **16 Comparing Observed and Expected Counts 299**

Education and Anomia	299
Observed and Expected Counts	300
The Chi-Square Statistic	302

College Degrees and Perception of Life 305

A Larger Table 306

A One-Sample Chi-Square Test 308

Power Concerns 310

Summary 311

What's Next? 311

Exercises 312

## **17 Nonparametric Tests 315**

Nonparametric Tests for Paired Data 316

Sign Test 316

McNemar's Test 318

Wilcoxon Test 320

Mann-Whitney Test 322

Kruskal-Wallis Test 324

Runs Test 325

Summary 327

How to Obtain Nonparametric Tests 327

Chi-Square Test 327

Binomial Test 329

Runs Test 330

Two-Independent-Samples Tests 331

Several-Independent-Samples Tests 332

Two-Related-Samples Tests 333

Options: Descriptive Statistics and Missing Values 335

Exercises 336

## **Part 4 Examining Relationships**

### **18 Measuring Association 339**

The Strength of a Relationship 340

Measures of Association 341

Measures of Association for Nominal Variables 342

Measures Based on Chi-Square 342

Proportional Reduction in Error 344

Measures of Association for Ordinal Variables 350

Concordant and Discordant Pairs 351

Measures Based on Concordant and Discordant Pairs 352

Correlation-Based Measures 355

Measuring Agreement 356

Summary 358

What's Next? 358

Exercises 359

### **19 Linear Regression and Correlation 363**

Life Expectancy and Birthrate 364

Choosing the "Best" Line 365

Calculating the Least-Squares Line 370

Calculating Predicted Values and Residuals 372

Determining How Well the Line Fits 373

Explaining Variability 377

Some Warnings 379

Summary 381

What's Next? 381

How to Obtain a Linear Regression 382

Statistics: Further Information on the Model 383

Residual Plots: Basic Residual Analysis 384

Linear Regression Save: Creating New Variables 386  
Linear Regression Options 388  
Exercises 389

## **20 Testing Regression Hypotheses 397**

The Population Regression Line 397  
Assumptions Needed for Testing Hypotheses 398  
Testing Hypotheses 399  
Testing that the Slope Is Zero 400  
Confidence Intervals for the Slope and Intercept 402  
Predicting Life Expectancy 402  
Predicting Means and Individual Observations 403  
Standard Error of the Predicted Mean 404  
Confidence Intervals for the Predicted Means 405  
Prediction Intervals for Individual Cases 407  
Summary 409  
What's Next? 409  
How to Obtain a Bivariate Correlation 409  
Options: Additional Statistics and Missing Data 411  
How to Obtain a Partial Correlation 412  
Options: Additional Statistics and Missing Data 413  
Exercises 414

## **21 Analyzing Residuals 421**

Residuals 422  
Standardized Residuals 423  
Studentized Residuals 424  
Checking for Normality 425  
Checking for Constant Variance 428  
Checking Linearity 430  
Checking Independence 433



A Final Comment on Assumptions 434

Looking for Influential Points 435

Studentized Deleted Residuals 437

Summary 438

What's Next? 438

Exercises 439

## **22 Building Multiple Regression Models 445**

Predicting Life Expectancy 446

The Model 446

Assumptions for Multiple Regression 447

Examining the Variables 448

Looking at How Well the Model Fits 450

Examining the Coefficients 452

Interpreting the Partial Regression Coefficients 454

Changing the Model 455

Partial Correlation Coefficients 456

Tolerance and Multicollinearity 457

Beta Coefficients 458

Building a Regression Model 459

Methods for Selecting Variables 460

Summary 468

What's Next? 468

How to Obtain a Multiple Linear Regression 469

Options: Variable Selection Criteria 470

Exercises 472

## **23 Multiple Regression Diagnostics 479**

Examining Normality 480

Scatterplots of Residuals 482