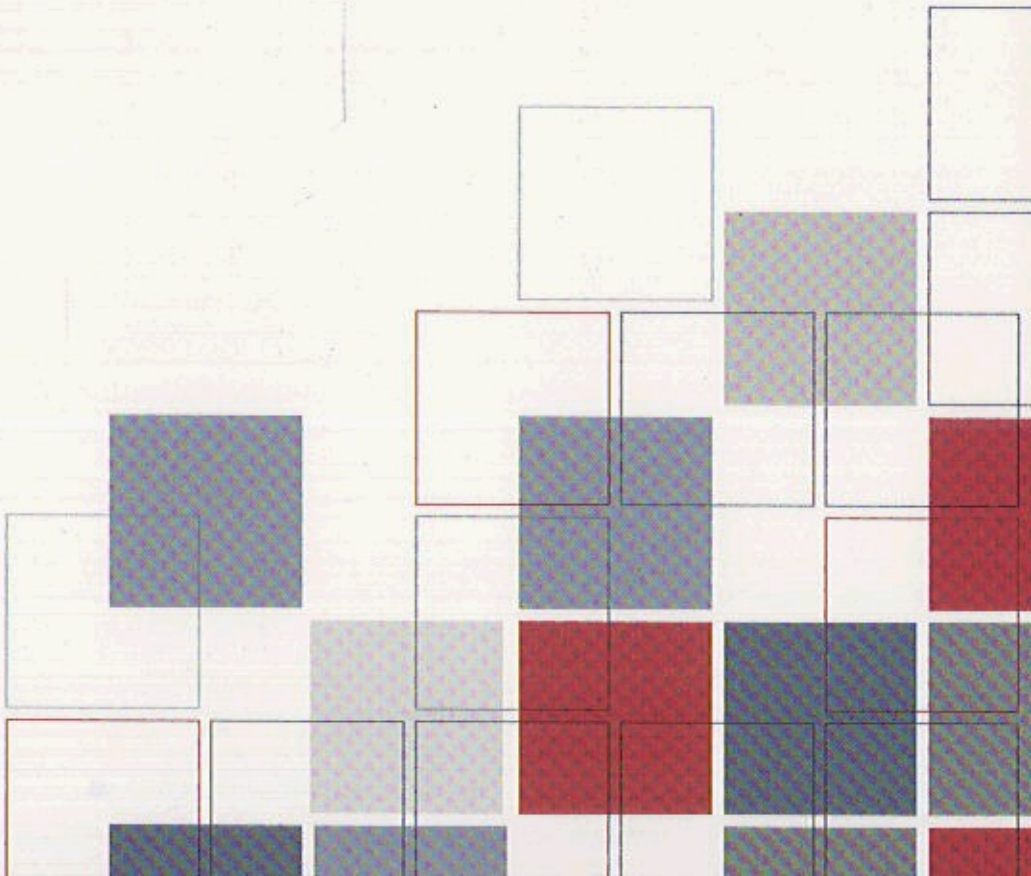


> SPSS 14.0 Guide to Data Analysis

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 of SPSS 5**

- Starting SPSS 5
 - Help Is Always at Hand 6
- Copying the Data Files 7
- Opening a Data File 8
- Statistical Procedures 11
 - The Viewer Window 14
 - Viewer Objects 17
- 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

3 **Sources of Data 31**

Know Your Data 31

Survey Data 31

Asking the Question 32

Measuring Time 33

Selecting Participants 34

Selecting a Sample 35

General Social Survey 37

Random-Digit Dialing 38

Internet Surveys 39

Designing Experiments 39

Random Assignment 40

Minimizing Bias 42

Summary 43

What's Next? 43

Exercises 44

Part 2 **Describing Data**

4 **Counting Responses 47**

Describing Variables 48

A Simple Frequency Table 49

Sorting Frequency Tables 53

Pie Charts 54

Bar Charts 56

Summarizing Internet Time 57

Histograms 59

Mode and Median 61

Percentiles 63

Summary 64

What's Next? 64

How to Obtain a Frequency Table	65
Format: Appearance of the Frequency Table	66
Statistics: Univariate Statistics	66
Charts: Bar Charts, Pie Charts, and Histograms	67
Exercises	68

5 **Computing Descriptive Statistics** 77

Summarizing Data	78
Scales of Measurement	78
Mode, Median, and Arithmetic Average	80
Comparing Mean and Median	81
Summarizing Time Spent Online	84
Measures of Variability	87
Range	88
Variance and Standard Deviation	88
The Coefficient of Variation	89
Standard Scores	90
Summary	92
What's Next?	92
How to Obtain Univariate Descriptive Statistics	93
Options: Choosing Statistics and Sorting Variables	94
Exercises	95

6 **Comparing Groups** 101

Age, Education, and Internet Use	102
Plotting Means	103
Layers: Defining Subgroups by More than One Variable	103
Summary	106
What's Next?	107
How to Obtain Subgroup Means	107
Layers: Defining Subgroups by More than One Variable	108
Options: Additional Statistics and Display of Labels	109
Exercises	110

7

Looking at Distributions 115

- Marathon Completion Times 116
 - Age and Gender 118
 - Marathon Times for Mature Runners 126
- Summary 129
- What's Next? 129
- How to Explore Distributions 129
 - Explore Statistics 131
 - Graphical Displays 131
 - Options 133
- Exercises 134

8

Counting Responses for Combinations of Variables 139

- Library Use and Education 140
 - Row and Column Percentages 142
 - Bar Charts 146
 - Adding Control Variables 148
 - Library Use and the Internet 150
- Summary 154
- What's Next? 154
- How to Obtain a Crosstabulation 155
 - Layers: Three or More Variables at Once 156
 - Cells: Percentages, Expected Counts, and Residuals 158
 - Bivariate Statistics 159
 - Format: Adjusting the Table Format 159
- Exercises 160

9

Plotting Data 169

- Examining Population Indicators 170
 - Simple Scatterplots 170
 - Scatterplot Matrices 174
 - Overlay Plots 176
 - Three-Dimensional Plots 179
 - Identifying Unusual Points 182
 - Rotating 3-D Scatterplots 183

Summary	184
What's Next?	184
How to Obtain a Scatterplot	184
Obtaining a Simple Scatterplot	185
Obtaining an Overlay Scatterplot	186
Obtaining a Scatterplot Matrix	188
Obtaining a 3-D Scatterplot	189
Editing a Scatterplot	190
Exercises	196

Part 3 **Testing Hypotheses**

10 Evaluating Results from Samples 201

From Sample to Population	202
A Computer Model	202
The Effect of Sample Size	206
The Binomial Test	209
Summary	212
What's Next?	212
Exercises	213

11 The Normal Distribution 215

The Normal Distribution	215
Samples from a Normal Distribution	219
Means from a Normal Population	219
Are the Sample Results Unlikely?	221
Testing a Hypothesis	223
Means from Non-Normal Distributions	224
Means from a Uniform Distribution	224
Summary	226
What's Next?	226
Exercises	227

12 **Testing a Hypothesis about a Single Mean 235**

Examining the Data 236

The T Distribution 238

 Calculating the T Statistic 240

Confidence Intervals 241

 Other Confidence Levels 244

 Confidence Interval for a Difference 245

 Confidence Intervals and Hypothesis Tests 245

Null Hypotheses and Alternative Hypotheses 246

 Rejecting the Null Hypothesis 247

Summary 248

What's Next? 248

How to Obtain a One-Sample T Test 249

 Options: Confidence Level and Missing Data 249

Exercises 250

13 **Testing a Hypothesis about Two Related Means 255**

Marathon Runners in Paired Designs 256

 Looking at Differences 257

 Is the Mean Difference Zero? 259

 Two Approaches 259

The Paired-Samples T Test 261

 Are You Positive? 262

 Some Possible Problems 263

 Examining Normality 263

Summary 265

What's Next? 266

How to Obtain a Paired-Samples T Test 266

 Options: Confidence Level and Missing Data 267

Exercises 268

14

Testing a Hypothesis about Two Independent Means 271

- Examining Television Viewing 272
 - Distribution of Differences 275
 - Standard Error of the Mean Difference 276
 - Computing the T Statistic 277
 - Output from the Two-Independent-Samples T Test 277
 - Confidence Intervals for the Mean Difference 279
 - Testing the Equality of Variances 279
- Effect of Outliers 281
- Introducing Education 283
 - Can You Prove the Null Hypothesis? 288
 - Interpreting the Observed Significance Level 289
 - Power 290
 - Monitoring Death Rates 290
 - Does Significant Mean Important? 293
- Summary 294
- What's Next? 294
- How to Obtain an Independent-Samples T Test 294
 - Define Groups: Specifying the Subgroups 295
 - Options: Confidence Level and Missing Data 296
- Exercises 297

15

One-Way Analysis of Variance 303

- Hours in a Work Week 304
 - Describing the Data 304
 - Confidence Intervals for the Group Means 305
 - Testing the Null Hypothesis 306
 - Assumptions Needed for Analysis of Variance 307
 - Analyzing the Variability 309
 - Comparing the Two Estimates of Variability 311
 - The Analysis-of-Variance Table 311
- Multiple Comparison Procedures 315
 - Television Viewing, Education, and Internet Use 318
- Summary 322

What's Next?	322
How to Obtain a One-Way Analysis of Variance	322
Post Hoc Multiple Comparisons: Finding the Difference	324
Options: Statistics and Missing Data	324
Exercises	326

16

Two-Way Analysis of Variance 331

The Design	332
Examining the Data	334
Testing Hypotheses	335
Degree and Gender Interaction	339
Necessary Assumptions	340
Analysis-of-Variance Table	341
Testing the Degree-by-Gender Interaction	342
Testing the Main Effects	343
Removing the Interaction Effect	344
Where Are the Differences?	346
Multiple Comparison Results	346
Checking Assumptions	347
A Look at Television	349
Extensions	351
Summary	351
What's Next?	351
How to Obtain a GLM Univariate Analysis	352
GLM Univariate: Model	353
GLM Univariate: Plots	354
GLM Univariate: Post Hoc	355
GLM Univariate: Options	355
GLM Univariate: Save	356
Exercises	357

17 **Comparing Observed and Expected Counts 363**

- Freedom or Manners? 364
 - Observed and Expected Counts 365
 - The Chi-Square Statistic 367
 - A Larger Table 371
- Does College Open Doors? 374
- A One-Sample Chi-Square Test 375
- Power Concerns 377
- Summary 378
- What's Next? 378
- Exercises 379

18 **Nonparametric Tests 383**

- Nonparametric Tests for Paired Data 384
 - Sign Test 387
 - Wilcoxon Test 390
 - Who's Sending E-mail? 392
- Mann-Whitney Test 394
- Kruskal-Wallis Test 396
- Friedman Test 397
- Summary 400
- How to Obtain Nonparametric Tests 400
 - Chi-Square Test 400
 - Binomial Test 402
 - Two-Independent-Samples Tests 403
 - Several-Independent-Samples Tests 404
 - Two-Related-Samples Tests 406
 - Several-Related-Samples Tests 407
 - Options: Descriptive Statistics and Missing Values 408
- Exercises 409

Part 4 **Examining Relationships**

19 **Measuring Association 413**

- Components of the Justice System 414
- Proportional Reduction in Error 417
- Measures of Association for Ordinal Variables 423
 - Concordant and Discordant Pairs 424
 - Measures Based on Concordant and Discordant Pairs 425
 - Evaluating the Components 428
 - Measuring Agreement 429
 - Correlation-Based Measures 431
- Measures Based on the Chi-Square Statistic 432
- Summary 436
- What's Next? 436
- Exercises 437

20 **Linear Regression and Correlation 441**

- Life Expectancy and Birthrate 442
 - Choosing the "Best" Line 443
- Calculating the Least-Squares Line 448
 - Calculating Predicted Values and Residuals 449
 - Determining How Well the Line Fits 450
 - Explaining Variability 454
 - Some Warnings 456
- Summary 457
- What's Next? 458
- How to Obtain a Linear Regression 458
 - Statistics: Further Information on the Model 460
 - Residual Plots: Basic Residual Analysis 461
 - Linear Regression Save: Creating New Variables 462
 - Linear Regression Options 465
- Exercises 467

21	Testing Regression Hypotheses	475
	The Population Regression Line	475
	Assumptions Needed for Testing Hypotheses	476
	Testing Hypotheses	477
	Testing that the Slope Is Zero	478
	Confidence Intervals for the Slope and Intercept	480
	Predicting Life Expectancy	480
	Predicting Means and Individual Observations	481
	Standard Error of the Predicted Mean	482
	Confidence Intervals for the Predicted Means	483
	Prediction Intervals for Individual Cases	484
	Summary	486
	What's Next?	486
	How to Obtain a Bivariate Correlation	486
	Options: Additional Statistics and Missing Data	488
	How to Obtain a Partial Correlation	488
	Options: Additional Statistics and Missing Data	490
	Exercises	491

22 Analyzing Residuals 497

Residuals	498
Standardized Residuals	499
Studentized Residuals	500
Checking for Normality	501
Checking for Constant Variance	504
Checking Linearity	506
Checking Independence	509
A Final Comment on Assumptions	510
Looking for Influential Points	510
Studentized Deleted Residuals	513
Summary	514
What's Next?	514
Exercises	514

23 **Building Multiple Regression Models 521**

- Predicting Life Expectancy 522
 - The Model 522
 - Assumptions for Multiple Regression 523
 - Examining the Variables 524
 - Looking at How Well the Model Fits 526
 - Examining the Coefficients 528
 - Interpreting the Partial Regression Coefficients 530
 - Changing the Model 531
 - Partial Correlation Coefficients 532
 - Tolerance and Multicollinearity 533
 - Beta Coefficients 534
- Building a Regression Model 535
 - Methods for Selecting Variables 536
- Summary 544
- What's Next? 544
- How to Obtain a Multiple Linear Regression 545
 - Options: Variable Selection Criteria 546
- Exercises 548

24 **Multiple Regression Diagnostics 557**

- Examining Normality 558
- Scatterplots of Residuals 560
- Leverage 563
- Changes in the Coefficients 564
- Cook's Distance 565
- Plots against Independent Variables 566
 - Partial Regression Plot 569
- Why Bother? 570
- Summary 570
- Exercises 571

Appendices

A **Obtaining Charts in SPSS 577**

- Overview 577
- Creating Bar Charts 578
 - Creating a Chart Comparing Groups of Cases 578
 - Data Summary Options 580
 - Creating a Chart Comparing Several Variables 581
 - Changing the Summary Statistic 583
 - Options in Creating Charts 584
- Modifying Charts 584
 - Modifying Chart Options 585
 - Hints on Editing Charts 586
 - Saving Chart Files 587
- Line and Area Charts 587
- Pie Charts 588
- Boxplots 588
 - Case Labels 589
- Error Bar Charts 589
- Histograms 590
- Normal Probability Plots 592

B **Transforming and Selecting Data 593**

- Data Transformations 593
 - Transformations at a Glance 594
 - Saving Changes 595
 - Delaying Processing of Transformations 595
 - Recoding Values 597
- Computing Variables 601
 - The Calculator Pad 602
 - Automatic Recoding 607
 - Conditional Transformations 608
- Case Selection 613
 - Temporary or Permanent Selection 614
 - Other Selection Methods 616

C	The T Distribution	618
D	Areas under the Normal Curve	620
E	Descriptions of Data Files	623
F	Answers to Selected Exercises	625
	Bibliography	639
	Index	641