Contents

Preface xiv

1 Introduction 1

1.1 Multivariate Statistics: Why? 1

1.1.1 The Domain of Multivariate Statistics: Numbers of IVs and DVs 2

1.1.2 Experimental and Nonexperimental Research 2

1.1.3 Computers and Multivariate Statistics 3

1.1.4 Garbage In, Roses Out? 4

1.2 Some Useful Definitions 5

1.2.1 Continuous, Discrete, and Dichotomous Data 5

1.2.2 Samples and Populations 6

1.2.3 Descriptive and Inferential Statistics 7

1.2.4 Orthogonality: Standard and Sequential Analyses 7

1.3 Linear Combinations of Variables 9

1.4 Number and Nature of Variables to Include 10

1.5 Statistical Power 10

1.6 Data Appropriate for Multivariate Statistics 11

1.6.1 The Data Matrix 11

1.6.2 The Correlation Matrix 12

1.6.3 The Variance–Covariance Matrix 12

1.6.4 The Sum-of-Squares and Cross-Products Matrix 13

1.6.5 Residuals 14

1.7 Organization of the Book 14

2 A Guide to Statistical Techniques: Using the Book 15

2.1 Research Questions and Associated Techniques 15

2.1.1 Degree of Relationship Among Variables 15

2.1.1.1 Bivariate r 16

2.1.1.2 Multiple R 16

2.1.1.3 Sequential R 16

2.1.1.4 Canonical R 16

2.1.1.5 Multiway Frequency Analysis 17

2.1.1.6 Multilevel Modeling 17

2.1.2 Significance of Group Differences 17

2.1.2.1 One-Way ANOVA and t Test 17

2.1.2.2 One-Way ANCOVA 17

2.1.2.3 Factorial ANOVA 18

2.1.2.4 Factorial ANCOVA 18

2.1.2.5 Hotelling’s $T^2$ 18

2.1.2.6 One-Way MANOVA 18

2.1.2.7 One-Way MANCOVA 19

2.1.2.8 Factorial MANOVA 19

2.1.2.9 Factorial MANCOVA 19

2.1.2.10 Profile Analysis of Repeated Measures 19

2.1.3 Prediction of Group Membership 20

2.1.3.1 One-Way Discriminant Analysis 20

2.1.3.2 Sequential One-Way Discriminant Analysis 20

2.1.3.3 Multiway Frequency Analysis (Logit) 21

2.1.3.4 Logistic Regression 21

2.1.3.5 Sequential Logistic Regression 21

2.1.3.6 Factorial Discriminant Analysis 21

2.1.3.7 Sequential Factorial Discriminant Analysis 22

2.1.4 Structure 22

2.1.4.1 Principal Components 22

2.1.4.2 Factor Analysis 22

2.1.4.3 Structural Equation Modeling 22

2.1.5 Time Course of Events 22

2.1.5.1 Survival/Failure Analysis 23

2.1.5.2 Time-Series Analysis 23

2.2 Some Further Comparisons 23

2.3 A Decision Tree 24

2.4 Technique Chapters 27

2.5 Preliminary Check of the Data 28

3 Review of Univariate and Bivariate Statistics 29

3.1 Hypothesis Testing 29

3.1.1 One-Sample $z$ Test as Prototype 30

3.1.2 Power 32

3.1.3 Extensions of the Model 32

3.1.4 Controversy Surrounding Significance Testing 33

3.2 Analysis of Variance 33

3.2.1 One-Way Between-Subjects ANOVA 34

3.2.2 Factorial Between-Subjects ANOVA 36

3.2.3 Within-Subjects ANOVA 38

3.2.4 Mixed Between-Within-Subjects ANOVA 40

3.2.5 Design Complexity 41

3.2.5.1 Nesting 41

3.2.5.2 Latin-Square Designs 42

3.2.5.3 Unequal $n$ and Nonorthogonality 42

3.2.5.4 Fixed and Random Effects 43

3.2.6 Specific Comparisons 43

3.2.6.1 Weighting Coefficients for Comparisons 43

3.2.6.2 Orthogonality of Weighting Coefficients 44

3.2.6.3 Obtained $F$ for Comparisons 44

3.2.6.4 Critical $F$ for Planned Comparisons 45

3.2.6.5 Critical $F$ for Post Hoc Comparisons 45

3.3 Parameter Estimation 46

3.4 Effect Size 47

A01_TABA0541_07_ALC_FM.indd   3
0x0 5/17/18   8:59 PM
4 Cleaning Up Your Act: Screening Data Prior to Analysis 52

4.1 Important Issues in Data Screening 53
  4.1.1 Accuracy of Data File 53
  4.1.2 Honest Correlations 53
    4.1.2.1 Inflated Correlation 53
    4.1.2.2 Deflated Correlation 53
  4.1.3 Missing Data 54
    4.1.3.1 Deleting Cases or Variables 57
    4.1.3.2 Estimating Missing Data 57
    4.1.3.3 Using a Missing Data Correlation Matrix 61
    4.1.3.4 Treating Missing Data as Data 61
    4.1.3.5 Choosing Among Methods for Dealing with Missing Data 62
  4.1.4 Outliers 62
    4.1.4.1 Detecting Univariate and Multivariate Outliers 63
    4.1.4.2 Describing Outliers 66
    4.1.4.3 Reducing the Influence of Outliers 66
    4.1.4.4 Outliers in a Solution 67
  4.1.5 Normality, Linearity, and Homoscedasticity 67
    4.1.5.1 Normality 68
    4.1.5.2 Linearity 72
    4.1.5.3 Homoscedasticity, Homogeneity of Variance, and Homogeneity of Variance-Covariance Matrices 73
  4.1.6 Common Data Transformations 75
  4.1.7 Multicollinearity and Singularity 76
  4.1.8 A Checklist and Some Practical Recommendations 79

4.2 Complete Examples of Data Screening 79
  4.2.1 Screening Ungrouped Data 80
    4.2.1.1 Accuracy of Input, Missing Data, Distributions, and Univariate Outliers 81
    4.2.1.2 Linearity and Homoscedasticity 84
    4.2.1.3 Transformation 84
    4.2.1.4 Detecting Multivariate Outliers 84
    4.2.1.5 Variables Causing Cases to Be Outliers 86
    4.2.1.6 Multicollinearity 88
  4.2.2 Screening Grouped Data 88
    4.2.2.1 Accuracy of Input, Missing Data, Distributions, Homogeneity of Variance, and Univariate Outliers 89
    4.2.2.2 Linearity 93
    4.2.2.3 Multivariate Outliers 93
    4.2.2.4 Variables Causing Cases to Be Outliers 94
    4.2.2.5 Multicollinearity 97

5 Multiple Regression 99
  5.1 General Purpose and Description 99
  5.2 Kinds of Research Questions 101
    5.2.1 Degree of Relationship 101
    5.2.2 Importance of IVs 102
    5.2.3 Adding IVs 102
    5.2.4 Changing IVs 102
    5.2.5 Contingencies Among IVs 102
    5.2.6 Comparing Sets of IVs 102
    5.2.7 Predicting DV Scores for Members of a New Sample 103
    5.2.8 Parameter Estimates 103
  5.3 Limitations to Regression Analyses 103
    5.3.1 Theoretical Issues 103
    5.3.2 Practical Issues 104
      5.3.2.1 Ratio of Cases to IVs 105
      5.3.2.2 Absence of Outliers Among the IVs and on the DV 105
      5.3.2.3 Absence of Multicollinearity and Singularity 106
      5.3.2.4 Normality, Linearity, and Homoscedasticity of Residuals 106
      5.3.2.5 Independence of Errors 108
      5.3.2.6 Absence of Outliers in the Solution 109
  5.4 Fundamental Equations for Multiple Regression 109
    5.4.1 General Linear Equations 110
    5.4.2 Matrix Equations 111
    5.4.3 Computer Analyses of Small-Sample Example 113
  5.5 Major Types of Multiple Regression 115
    5.5.1 Standard Multiple Regression 115
    5.5.2 Sequential Multiple Regression 116
    5.5.3 Statistical (Stepwise) Regression 117
    5.5.4 Choosing Among Regression Strategies 121
  5.6 Some Important Issues 121
    5.6.1 Importance of IVs 121
      5.6.1.1 Standard Multiple Regression 122
      5.6.1.2 Sequential or Statistical Regression 123
      5.6.1.3 Commonality Analysis 123
      5.6.1.4 Relative Importance Analysis 125
    5.6.2 Statistical Inference 128
      5.6.2.1 Test for Multiple R 128
      5.6.2.2 Test of Regression Components 129
      5.6.2.3 Test of Added Subset of IVs 130
      5.6.2.4 Confidence Limits 130
      5.6.2.5 Comparing Two Sets of Predictors 131
    5.6.3 Adjustment of $R^2$ 132
    5.6.4 Suppressor Variables 133
    5.6.5 Regression Approach to ANOVA 134
    5.6.6 Centering When Interactions and Powers of IVs Are Included 135
    5.6.7 Mediation in Causal Sequence 137
5.7 Complete Examples of Regression Analysis 138
5.7.1 Evaluation of Assumptions 139
  5.7.1.1 Ratio of Cases to IVs 139
  5.7.1.2 Normality, Linearity, Homoscedasticity, and Independence of Residuals 139
  5.7.1.3 Outliers 142
  5.7.1.4 Multicollinearity and Singularity 144
5.7.2 Standard Multiple Regression 144
5.7.3 Sequential Regression 150
5.7.4 Example of Standard Multiple Regression with Missing Values Multiply Imputed 154

5.8 Comparison of Programs 162
  5.8.1 IBM SPSS Package 163
  5.8.2 SAS System 165
  5.8.3 SYSTAT System 166

6 Analysis of Covariance 167
  6.1 General Purpose and Description 167
  6.2 Kinds of Research Questions 170
  6.2.1 Main Effects of IVs 170
  6.2.2 Interactions Among IVs 170
  6.2.3 Specific Comparisons and Trend Analysis 170
  6.2.4 Effects of Covariates 170
  6.2.5 Effect Size 171
  6.2.6 Parameter Estimates 171

6.3 Limitations to Analysis of Covariance 171
  6.3.1 Theoretical Issues 171
  6.3.2 Practical Issues 172
  6.3.2.1 Unequal Sample Sizes, Missing Data, and Ratio of Cases to IVs 172
  6.3.2.2 Absence of Outliers 172
  6.3.2.3 Absence of Multicollinearity and Singularity 172
  6.3.2.4 Normality of Sampling Distributions 173
  6.3.2.5 Homogeneity of Variance 173
  6.3.2.6 Linearity 173
  6.3.2.7 Homogeneity of Regression 173
  6.3.2.8 Reliability of Covariates 174

6.4 Fundamental Equations for Analysis of Covariance 174
  6.4.1 Sums of Squares and Cross-Products 175
  6.4.2 Significance Test and Effect Size 177
  6.4.3 Computer Analyses of Small-Sample Example 178

6.5 Some Important Issues 179
  6.5.1 Choosing Covariates 179
  6.5.2 Evaluation of Covariates 180
  6.5.3 Test for Homogeneity of Regression 180
  6.5.4 Design Complexity 181
    6.5.4.1 Within-Subjects and Mixed Within-Between Designs 181
    6.5.4.2 Unequal Sample Sizes 182

6.6 Complete Example of Analysis of Covariance 189
  6.6.1 Evaluation of Assumptions 189
    6.6.1.1 Unequal n and Missing Data 189
    6.6.1.2 Normality 191
    6.6.1.3 Linearity 191
    6.6.1.4 Outliers 191
    6.6.1.5 Multicollinearity and Singularity 192
    6.6.1.6 Homogeneity of Variance 192
    6.6.1.7 Homogeneity of Regression 193
    6.6.1.8 Reliability of Covariates 193
  6.6.2 Analysis of Covariance 193
    6.6.2.1 Main Analysis 193
    6.6.2.2 Evaluation of Covariates 196
    6.6.2.3 Homogeneity of Regression Run 197

6.7 Comparison of Programs 200
  6.7.1 IBM SPSS Package 200
  6.7.2 SAS System 200
  6.7.3 SYSTAT System 200

7 Multivariate Analysis of Variance and Covariance 203
  7.1 General Purpose and Description 203
  7.2 Kinds of Research Questions 206
  7.2.1 Main Effects of IVs 206
  7.2.2 Interactions Among IVs 207
  7.2.3 Importance of DVs 207
  7.2.4 Parameter Estimates 207
  7.2.5 Specific Comparisons and Trend Analysis 207
  7.2.6 Effect Size 208
  7.2.7 Effects of Covariates 208
  7.2.8 Repeated-Measures Analysis of Variance 208

7.3 Limitations to Multivariate Analysis of Variance and Covariance 208
  7.3.1 Theoretical Issues 208
  7.3.2 Practical Issues 209
    7.3.2.1 Unequal Sample Sizes, Missing Data, and Power 209
    7.3.2.2 Multivariate Normality 210
    7.3.2.3 Absence of Outliers 210
    7.3.2.4 Homogeneity of Variance–Covariance Matrices 210
    7.3.2.5 Linearity 211
    7.3.2.6 Homogeneity of Regression 211
    7.3.2.7 Reliability of Covariates 211
    7.3.2.8 Absence of Multicollinearity and Singularity 211

7.4 Fundamental Equations for Multivariate Analysis of Variance and Covariance 212
  7.4.1 Multivariate Analysis of Variance 212
10.6.8 Importance of Predictors 373
10.6.9 Logistic Regression for Matched Groups 374
10.7 Complete Examples of Logistic Regression 374
10.7.1 Evaluation of Limitations 374
10.7.1.1 Ratio of Cases to Variables and Missing Data 374
10.7.1.2 Multicollinearity 376
10.7.1.3 Outliers in the Solution 376
10.7.2 Direct Logistic Regression with Two-Category Outcome and Continuous Predictors 377
10.7.2.1 Limitation: Linearity in the Logit 377
10.7.2.2 Direct Logistic Regression with Two-Category Outcome 377
10.7.3 Sequential Logistic Regression with Three Categories of Outcome 384
10.7.3.1 Limitations of Multinomial Logistic Regression 384
10.7.3.2 Sequential Multinomial Logistic Regression 387
10.8 Comparison of Programs 396
10.8.1 IBM SPSS Package 396
10.8.2 SAS System 399
10.8.3 SYSTAT System 400
11 Survival/Failure Analysis 401
11.1 General Purpose and Description 401
11.2 Kinds of Research Questions 403
11.2.1 Proportions Surviving at Various Times 403
11.2.2 Group Differences in Survival 403
11.2.3 Survival Time with Covariates 403
11.2.3.1 Treatment Effects 403
11.2.3.2 Importance of Covariates 403
11.2.3.3 Parameter Estimates 404
11.2.3.4 Contingencies Among Covariates 404
11.2.3.5 Effect Size and Power 404
11.3 Limitations to Survival Analysis 404
11.3.1 Theoretical Issues 404
11.3.2 Practical Issues 404
11.3.2.1 Sample Size and Missing Data 404
11.3.2.2 Normality of Sampling Distributions, Linearity, and Homoscedasticity 405
11.3.2.3 Absence of Outliers 405
11.3.2.4 Differences Between Withdrawn and Remaining Cases 405
11.3.2.5 Change in Survival Conditions over Time 405
11.3.2.6 Proportionality of Hazards 405
11.3.2.7 Absence of Multicollinearity 405
11.4 Fundamental Equations for Survival Analysis 405
11.4.1 Life Tables 406
11.4.2 Standard Error of Cumulative Proportion Surviving 408
11.4.3 Hazard and Density Functions 408
11.4.4 Plot of Life Tables 409
11.4.5 Test for Group Differences 410
11.4.6 Computer Analyses of Small-Sample Example 411
11.5 Types of Survival Analyses 415
11.5.1 Actuarial and Product-Limit Life Tables and Survivor Functions 415
11.5.2 Prediction of Group Survival Times from Covariates 417
11.5.2.1 Direct, Sequential, and Statistical Analysis 417
11.5.2.2 Cox Proportional-Hazards Model 417
11.5.2.3 Accelerated Failure-Time Models 419
11.5.2.4 Choosing a Method 423
11.6 Some Important Issues 423
11.6.1 Proportionality of Hazards 423
11.6.2 Censored Data 424
11.6.2.1 Right-Censored Data 425
11.6.2.2 Other Forms of Censoring 425
11.6.3 Effect Size and Power 425
11.6.4 Statistical Criteria 426
11.6.4.1 Test Statistics for Group Differences in Survival Functions 426
11.6.4.2 Test Statistics for Prediction from Covariates 427
11.6.5 Predicting Survival Rate 427
11.6.5.1 Regression Coefficients (Parameter Estimates) 427
11.6.5.2 Hazard Ratios 427
11.6.5.3 Expected Survival Rates 428
11.7 Complete Example of Survival Analysis 429
11.7.1 Evaluation of Assumptions 430
11.7.1.1 Accuracy of Input, Adequacy of Sample Size, Missing Data, and Distributions 430
11.7.1.2 Outliers 430
11.7.1.3 Differences Between Withdrawn and Remaining Cases 433
11.7.1.4 Change in Survival Experience over Time 433
11.7.1.5 Proportionality of Hazards 433
11.7.1.6 Multicollinearity 434
11.7.2 Cox Regression Survival Analysis 436
11.7.2.1 Effect of Drug Treatment 436
11.7.2.2 Evaluation of Other Covariates 436
11.8 Comparison of Programs 440
11.8.1 SAS System 444
11.8.2 IBM SPSS Package 445
11.8.3 SYSTAT System 445
12 Canonical Correlation 446
12.1 General Purpose and Description 446
12.2 Kinds of Research Questions 448
12.2.1 Number of Canonical Variate Pairs 448
12.2.2 Interpretation of Canonical Variates 448
12.2.3 Importance of Canonical Variates and Predictors 448
12.2.4 Canonical Variate Scores 449

12.3 Limitations 449
12.3.1 Theoretical Limitations 449
12.3.2 Practical Issues 450
12.3.2.1 Ratio of Cases to IVs 450
12.3.2.2 Normality, Linearity, and Homoscedasticity 450
12.3.2.3 Missing Data 451
12.3.2.4 Absence of Outliers 451
12.3.2.5 Absence of Multicollinearity and Singularity 451

12.4 Fundamental Equations for Canonical Correlation 451
12.4.1 Eigenvalues and Eigenvectors 452
12.4.2 Matrix Equations 454
12.4.3 Proportions of Variance Extracted 457
12.4.4 Computer Analyses of Small-Sample Example 458

12.5 Some Important Issues 462
12.5.1 Importance of Canonical Variates 462
12.5.2 Interpretation of Canonical Variates 463

12.6 Complete Example of Canonical Correlation 463
12.6.1 Evaluation of Assumptions 463
12.6.1.1 Missing Data 463
12.6.1.2 Normality, Linearity, and Homoscedasticity 463
12.6.1.3 Outliers 466
12.6.1.4 Multicollinearity and Singularity 467
12.6.2 Canonical Correlation 467

12.7 Comparison of Programs 473
12.7.1 SAS System 473
12.7.2 IBM SPSS Package 474
12.7.3 SYSTAT System 475

13 Principal Components and Factor Analysis 476
13.1 General Purpose and Description 476
13.2 Kinds of Research Questions 479
13.2.1 Number of Factors 479
13.2.2 Nature of Factors 479
13.2.3 Importance of Solutions and Factors 480
13.2.4 Testing Theory in FA 480
13.2.5 Estimating Scores on Factors 480

13.3 Limitations 480
13.3.1 Theoretical Issues 480
13.3.2 Practical Issues 481
13.3.2.1 Sample Size and Missing Data 481
13.3.2.2 Normality 482
13.3.2.3 Linearity 482
13.3.2.4 Absence of Outliers Among Cases 482

13.3.2.5 Absence of Multicollinearity and Singularity 482
13.3.2.6 Factorability of R 482
13.3.2.7 Absence of Outliers Among Variables 483

13.4 Fundamental Equations for Factor Analysis 483
13.4.1 Extraction 485
13.4.2 Orthogonal Rotation 487
13.4.3 Communalities, Variance, and Covariance 488
13.4.4 Factor Scores 489
13.4.5 Oblique Rotation 491
13.4.6 Computer Analyses of Small-Sample Example 493

13.5 Major Types of Factor Analyses 496
13.5.1 Factor Extraction Techniques 496
13.5.1.1 PCA Versus FA 496
13.5.1.2 Principal Components 498
13.5.1.3 Principal Factors 498
13.5.1.4 Image Factor Extraction 498
13.5.1.5 Maximum Likelihood Factor Extraction 499
13.5.1.6 Unweighted Least Squares Factoring 499
13.5.1.7 Generalized (Weighted) Least Squares Factoring 499
13.5.1.8 Alpha Factoring 499
13.5.2 Rotation 500
13.5.2.1 Orthogonal Rotation 500
13.5.2.2 Oblique Rotation 501
13.5.2.3 Geometric Interpretation 502
13.5.3 Some Practical Recommendations 503

13.6 Some Important Issues 504
13.6.1 Estimates of Communalities 504
13.6.2 Adequacy of Extraction and Number of Factors 504
13.6.3 Adequacy of Rotation and Simple Structure 507
13.6.4 Importance and Internal Consistency of Factors 508
13.6.5 Interpretation of Factors 509
13.6.6 Factor Scores 510
13.6.7 Comparisons Among Solutions and Groups 511

13.7 Complete Example of FA 511
13.7.1 Evaluation of Limitations 511
13.7.2 Principal Factors Extraction with Varimax Rotation 515
14 Structural Equation Modeling by Jodie B. Ullman 528

14.1 General Purpose and Description 528

14.2 Kinds of Research Questions 531

14.3 Limitations to Structural Equation Modeling 533

14.4 Fundamental Equations for Structural Equation Modeling 535

14.5 Some Important Issues 555

14.6 Complete Examples of Structural Equation Modeling Analysis 574

14.7 Comparison of Programs 607

15 Multilevel Linear Modeling 613

15.1 General Purpose and Description 613

15.2 Kinds of Research Questions 616

15.3 Assessing the Fit of the Model 560
15.3 Limitations to Multilevel Linear Modeling 618
  15.3.1 Theoretical Issues 618
  15.3.2 Practical Issues 618
    15.3.2.1 Sample Size, Unequal-n, and Missing Data 619
    15.3.2.2 Independence of Errors 619
    15.3.2.3 Absence of Multicollinearity and Singularity 620
  15.4 Fundamental Equations 620
    15.4.1 Intercepts-Only Model 623
      15.4.1.1 The Intercepts-Only Model: Level-1 Equation 623
      15.4.1.2 The Intercepts-Only Model: Level-2 Equation 623
      15.4.1.3 Computer Analyses of Intercepts-Only Model 624
    15.4.2 Model with a First-Level Predictor 627
      15.4.2.1 Level-1 Equation for a Model with a Level-1 Predictor 627
      15.4.2.2 Level-2 Equations for a Model with a Level-1 Predictor 628
      15.4.2.3 Computer Analysis of a Model with a Level-1 Predictor 630
    15.4.3 Model with Predictors at First and Second Levels 633
      15.4.3.1 Level-1 Equation for a Model with Predictors at Both Levels 633
      15.4.3.2 Level-2 Equations for a Model with Predictors at Both Levels 633
      15.4.3.3 Computer Analyses of Model with Predictors at First and Second Levels 634
  15.5 Types of MLM 638
    15.5.1 Repeated Measures 638
    15.5.2 Higher-Order MLM 642
    15.5.3 Latent Variables 642
    15.5.4 Nonnormal Outcome Variables 643
    15.5.5 Multiple Response Models 644
  15.6 Some Important Issues 644
    15.6.1 Intraclass Correlation 644
    15.6.2 Centering Predictors and Changes in Their Interpretations 646
    15.6.3 Interactions 648
    15.6.4 Random and Fixed Intercepts and Slopes 648
    15.6.5 Statistical Inference 651
      15.6.5.1 Assessing Models 651
      15.6.5.2 Tests of Individual Effects 652
    15.6.6 Effect Size 653
    15.6.7 Estimation Techniques and Convergence Problems 653
    15.6.8 Exploratory Model Building 654
  15.7 Complete Example of MLM 655
    15.7.1 Evaluation of Assumptions 656
  15.7.1.1 Sample Sizes, Missing Data, and Distributions 656
  15.7.1.2 Outliers 659
  15.7.1.3 Multicollinearity and Singularity 659
  15.7.1.4 Independence of Errors: Intraclass Correlations 659
  15.7.2 Multilevel Modeling 661
  15.8 Comparison of Programs 668
    15.8.1 SAS System 668
    15.8.2 IBM SPSS Package 670
    15.8.3 HLM Program 671
    15.8.4 MLwiN Program 671
    15.8.5 SYSTAT System 671

16 Multiway Frequency Analysis 672
  16.1 General Purpose and Description 672
  16.2 Kinds of Research Questions 673
    16.2.1 Associations Among Variables 673
    16.2.2 Effect on a Dependent Variable 674
    16.2.3 Parameter Estimates 674
    16.2.4 Importance of Effects 674
    16.2.5 Effect Size 674
    16.2.6 Specific Comparisons and Trend Analysis 674
  16.3 Limitations to Multiway Frequency Analysis 675
    16.3.1 Theoretical Issues 675
    16.3.2 Practical Issues 675
      16.3.2.1 Independence 675
      16.3.2.2 Ratio of Cases to Variables 675
      16.3.2.3 Adequacy of Expected Frequencies 675
      16.3.2.4 Absence of Outliers in the Solution 676
  16.4 Fundamental Equations for Multiway Frequency Analysis 676
    16.4.1 Screening for Effects 678
      16.4.1.1 Total Effect 678
      16.4.1.2 First-Order Effects 679
      16.4.1.3 Second-Order Effects 679
      16.4.1.4 Third-Order Effect 683
    16.4.2 Modeling 683
    16.4.3 Evaluation and Interpretation 685
      16.4.3.1 Residuals 685
      16.4.3.2 Parameter Estimates 686
    16.4.4 Computer Analyses of Small-Sample Example 690
  16.5 Some Important Issues 695
    16.5.1 Hierarchical and Nonhierarchical Models 695
    16.5.2 Statistical Criteria 696
      16.5.2.1 Tests of Models 696
      16.5.2.2 Tests of Individual Effects 696
    16.5.3 Strategies for Choosing a Model 696
      16.5.3.1 IBM SPSS HILOGLINEAR ( Hierarchical) 697
16.5.3.2 IBM SPSS GENLOG (General Log-Linear) 697
16.5.3.3 SAS CATMOD and IBM SPSS LOGLINEAR (General Log-Linear) 697

16.6 Complete Example of Multiway Frequency Analysis 698
16.6.1 Evaluation of Assumptions: Adequacy of Expected Frequencies 698
16.6.2 Hierarchical Log-Linear Analysis 700
16.6.2.1 Preliminary Model Screening 700
16.6.2.2 Stepwise Model Selection 702
16.6.2.3 Adequacy of Fit 702
16.6.2.4 Interpretation of the Selected Model 705

16.7 Comparison of Programs 710
16.7.1 IBM SPSS Package 710
16.7.2 SAS System 712
16.7.3 SYSTAT System 713

17 Time-Series Analysis 714
17.1 General Purpose and Description 714
17.2 Kinds of Research Questions 716
17.2.1 Pattern of Autocorrelation 717
17.2.2 Seasonal Cycles and Trends 717
17.2.3 Forecasting 717
17.2.4 Effect of an Intervention 718
17.2.5 Comparing Time Series 718
17.2.6 Time Series with Covariates 718
17.2.7 Effect Size and Power 718

17.3 Assumptions of Time-Series Analysis 718
17.3.1 Theoretical Issues 718
17.3.2 Practical Issues 718
17.3.2.1 Normality of Distributions of Residuals 719
17.3.2.2 Homogeneity of Variance and Zero Mean of Residuals 719
17.3.2.3 Independence of Residuals 719
17.3.2.4 Absence of Outliers 719
17.3.2.5 Sample Size and Missing Data 719

17.4 Fundamental Equations for Time-Series ARIMA Models 720
17.4.1 Identification of ARIMA (p, d, q) Models 720
17.4.1.1 Trend Components, d: Making the Process Stationary 721
17.4.1.2 Auto-Regressive Components 722
17.4.1.3 Moving Average Components 724
17.4.1.4 Mixed Models 724
17.4.1.5 ACFs and PACFs 724
17.4.2 Estimating Model Parameters 729
17.4.3 Diagnosing a Model 729
17.4.4 Computer Analysis of Small-Sample Time-Series Example 734

17.5 Types of Time-Series Analyses 737
17.5.1 Models with Seasonal Components 737
17.5.2 Models with Interventions 738
17.5.2.1 Abrupt, Permanent Effects 741
17.5.2.2 Abrupt, Temporary Effects 742
17.5.2.3 Gradual, Permanent Effects 745
17.5.2.4 Models with Multiple Interventions 746
17.5.3 Adding Continuous Variables 747

17.6 Some Important Issues 748
17.6.1 Patterns of ACFs and PACFs 748
17.6.2 Effect Size 751
17.6.3 Forecasting 752
17.6.4 Statistical Methods for Comparing Two Models 752

17.7 Complete Examples of Time-Series Analysis 753
17.7.1 Time-Series Analysis of Introduction of Seat Belt Law 753
17.7.1.1 Evaluation of Assumptions 754
17.7.1.2 Baseline Model Identification and Estimation 755
17.7.1.3 Baseline Model Diagnosis 758
17.7.1.4 Intervention Analysis 758
17.7.2 Time-Series Analysis of Introduction of a Dashboard to an Educational Computer Game 762
17.7.2.1 Evaluation of Assumptions 763
17.7.2.2 Baseline Model Identification and Diagnosis 765
17.7.2.3 Intervention Analysis 766

17.8 Comparison of Programs 771
17.8.1 IBM SPSS Package 771
17.8.2 SAS System 774
17.8.3 SYSTAT System 774

18 An Overview of the General Linear Model 775
18.1 Linearity and the General Linear Model 775
18.2 Bivariate to Multivariate Statistics and Overview of Techniques 775
18.2.1 Bivariate Form 775
18.2.2 Simple Multivariate Form 777
18.2.3 Full Multivariate Form 778
18.3 Alternative Research Strategies 782

Appendix A
A Skimpy Introduction to Matrix Algebra 783
A.1 The Trace of a Matrix 784
A.2 Addition or Subtraction of a Constant to a Matrix 784
A.3 Multiplication or Division of a Matrix by a Constant 784
A.4 Addition and Subtraction of Two Matrices 785
A.5 Multiplication, Transposes, and Square Roots of Matrice 785
A.6  Matrix “Division” (Inverses and Determinants) 786
A.7  Eigenvalues and Eigenvectors: Procedures for Consolidating Variance from a Matrix 788

Appendix B
Research Designs for Complete Examples 791
B.1  Women’s Health and Drug Study 791
B.2  Sexual Attraction Study 793
B.3  Learning Disabilities Data Bank 794
B.4  Reaction Time to Identify Figures 794
B.5  Field Studies of Noise-Induced Sleep Disturbance 795
B.6  Clinical Trial for Primary Biliary Cirrhosis 795
B.7  Impact of Seat Belt Law 795
B.8  The Selene Online Educational Game 796

Appendix C
Statistical Tables 797
C.1  Normal Curve Areas 798
C.2  Critical Values of the t Distribution for \( \alpha = .05 \) and .01, Two-Tailed Test 799
C.3  Critical Values of the F Distribution 800
C.4  Critical Values of Chi Square (\( \chi^2 \)) 804
C.5  Critical Values for Squares Multiple Correlation (\( R^2 \)) in Forward Stepwise Selection: \( \alpha = .05 \) 805
C.6  Critical Values for \( F_{MAX} (S_{MAX}^2 / S_{MIN}^2) \) Distribution for \( \alpha = .05 \) and .01 807

References 808
Index 815
Preface

Some good things seem to go on forever: friendship and updating this book. It is diffi-
cult to believe that the first edition manuscript was typewritten, with real cutting and
pasting. The publisher required a paper manuscript with numbered pages—that was
almost our downfall. We could write a book on multivariate statistics, but we couldn’t get the
same number of pages (about 1200, double-spaced) twice in a row. SPSS was in release 9.0,
and the other program we demonstrated was BMDP. There were a mere 11 chapters, of which
6 of them were describing techniques. Multilevel and structural equation modeling were not
yet ready for prime time. Logistic regression and survival analysis were not yet popular.

Material new to this edition includes a redo of all SAS examples, with a pretty new output
format and replacement of interactive analyses that are no longer available. We’ve also re-run
the IBM SPSS examples to show the new output format. We’ve tried to update the references in
all chapters, including only classic citations if they date prior to 2000. New work on relative im-
portance has been incorporated in multiple regression, canonical correlation, and logistic regress-
ion analysis—complete with demonstrations. Multiple imputation procedures for dealing with
missing data have been updated, and we’ve added a new time-series example, taking advantage
of an IBM SPSS expert modeler that replaces previous tea-leaf reading aspects of the analysis.

Our goals in writing the book remain the same as in all previous editions—to present com-
plex statistical procedures in a way that is maximally useful and accessible to researchers who
are not necessarily statisticians. We strive to be short on theory but long on conceptual under-
standing. The statistical packages have become increasingly easy to use, making it all the more
critical to make sure that they are applied with a good understanding of what they can and
cannot do. But above all else—what does it all mean?

We have not changed the basic format underlying all of the technique chapters, now 14 of
them. We start with an overview of the technique, followed by the types of research questions
the techniques are designed to answer. We then provide the cautionary tale—what you need to
worry about and how to deal with those worries. Then come the fundamental equations underly-
ing the technique, which some readers truly enjoy working through (we know because they help-
fully point out any errors and/or inconsistencies they find); but other readers discover they can
skim (or skip) the section without any loss to their ability to conduct meaningful analysis of their
research. The fundamental equations are in the context of a small, made-up, usually silly data set
for which computer analyses are provided—usually IBM SPSS and SAS. Next, we delve into is-
sues surrounding the technique (such as different types of the analysis, follow-up procedures to
the main analysis, and effect size, if it is not amply covered elsewhere). Finally, we provide one or
two full-bore analyses of an actual real-life data set together with a Results section appropriate for
a journal. Data sets for these examples are available at www.pearsonhighered.com in IBM SPSS,
SAS, and ASCII formats. We end each technique chapter with a comparison of features available
in IBM SPSS, SAS, SYSTAT and sometimes other specialized programs. SYSTAT is a statistical
package that we reluctantly had to drop a few editions ago for lack of space.

We apologize in advance for the heft of the book; it is not our intention to line the cof-
fers of chiropractors, physical therapists, acupuncturists, and the like, but there’s really just so
much to say. As to our friendship, it’s still going strong despite living in different cities. Art has
taken the place of creating belly dance costumes for both of us, but we remain silly in outlook,
although serious in our analysis of research.

The lineup of people to thank grows with each edition, far too extensive to list: students,
reviewers, editors, and readers who send us corrections and point out areas of confusion. As
always, we take full responsibility for remaining errors and lack of clarity.

Barbara G. Tabachnick
Linda S. Fidell
To locate your representative, use the Rep Locator tool. These are data files for the complete examples of Chapters 4 through 16 and 18, *Using Multivariate Statistics*, 6th edition. File names are referred to generically in the book, for example, SCREEN.*, and files are in two formats. Formats are as follows