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 Multway 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 T2 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
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
    6.5.5 Effect Size 187
    6.5.6 Parameter Estimates 187

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
    7.4.2 Multivariate Analysis of Covariance 212
7.4.2 Computer Analyses of Small-Sample Example 218
7.4.3 Multivariate Analysis of Covariance 221
7.5 Some Important Issues 223
7.5.1 MANOVA Versus ANOVAs 223
7.5.2 Criteria for Statistical Inference 223
7.5.3 Assessing DVs 224
7.5.3.1 Univariate F 224
7.5.3.2 Roy–Bargmann Stepdown Analysis 226
7.5.3.3 Using Discriminant Analysis 226
7.5.3.4 Choosing Among Strategies for Assessing DVs 227
7.5.4 Specific Comparisons and Trend Analysis 227
7.5.5 Design Complexity 228
7.5.5.1 Within-Subjects and Between-Within Designs 228
7.5.5.2 Unequal Sample Sizes 228
7.6 Complete Examples of Multivariate Analysis of Variance and Covariance 230
7.6.1 Evaluation of Assumptions 230
7.6.1.1 Unequal Sample Sizes and Missing Data 230
7.6.1.2 Multivariate Normality 231
7.6.1.3 Linearity 231
7.6.1.4 Outliers 232
7.6.1.5 Homogeneity of Variance–Covariance Matrices 233
7.6.1.6 Homogeneity of Regression 233
7.6.1.7 Reliability of Covariates 235
7.6.1.8 Multicollinearity and Singularity 235
7.6.2 Multivariate Analysis of Variance 235
7.6.3 Multivariate Analysis of Covariance 244
7.6.3.1 Assessing Covariates 244
7.6.3.2 Assessing DVs 245
7.7 Comparison of Programs 252
7.7.1 IBM SPSS Package 252
7.7.2 SAS System 254
7.7.3 SYSTAT System 255
8 Profile Analysis: The Multivariate Approach to Repeated Measures 256
8.1 General Purpose and Description 256
8.2 Kinds of Research Questions 257
8.2.1 Parallelism of Profiles 258
8.2.2 Overall Difference Among Groups 258
8.2.3 Flatness of Profiles 258
8.2.4 Contrasts Following Profile Analysis 258
8.2.5 Parameter Estimates 258
8.2.6 Effect Size 259
8.3 Limitations to Profile Analysis 259
8.3.1 Theoretical Issues 259
8.3.2 Practical Issues 259
8.3.2.1 Sample Size, Missing Data, and Power 259
8.3.2.2 Multivariate Normality 260
8.3.2.3 Absence of Outliers 260
8.3.2.4 Homogeneity of Variance–Covariance Matrices 260
8.3.2.5 Linearity 260
8.3.2.6 Absence of Multicollinearity and Singularity 260
8.4 Fundamental Equations for Profile Analysis 260
8.4.1 Differences in Levels 262
8.4.2 Parallelism 262
8.4.3 Flatness 265
8.4.4 Computer Analyses of Small-Sample Example 266
8.5 Some Important Issues 269
8.5.1 Univariate Versus Multivariate Approach to Repeated Measures 269
8.5.2 Contrasts in Profile Analysis 270
8.5.2.1 Parallelism and Flatness Significant, Levels Not Significant (Simple-Effects Analysis) 272
8.5.2.2 Parallelism and Levels Significant, Flatness Not Significant (Simple-Effects Analysis) 274
8.5.2.3 Parallelism, Levels, and Flatness Significant (Interaction Contrasts) 275
8.5.2.4 Only Parallelism Significant 276
8.5.3 Doubly Multivariate Designs 277
8.5.4 Classifying Profiles 279
8.5.5 Imputation of Missing Values 279
8.6 Complete Examples of Profile Analysis 280
8.6.1 Profile Analysis of Subscales of the WISC 280
8.6.1.1 Evaluation of Assumptions 280
8.6.1.2 Profile Analysis 283
8.6.2 Doubly Multivariate Analysis of Reaction Time 288
8.6.2.1 Evaluation of Assumptions 289
8.6.2.2 Doubly Multivariate Analysis of Slope and Intercept 290
8.7 Comparison of Programs 297
8.7.1 IBM SPSS Package 297
8.7.2 SAS System 298
8.7.3 SYSTAT System 298
9 Discriminant Analysis 299
9.1 General Purpose and Description 299
9.2 Kinds of Research Questions 302
9.2.1 Significance of Prediction 302
9.2.2 Number of Significant Discriminant Functions 302
9.2.3 Dimensions of Discrimination 302
9.2.4 Classification Functions 303
9.2.5 Adequacy of Classification 303
9.2.6 Effect Size 303
9.2.7 Importance of Predictor Variables 303
9.2.8 Significance of Prediction with Covariates 304
9.2.9 Estimation of Group Means 304
9.3 Limitations to Discriminant Analysis  
9.3.1 Theoretical Issues 304  
9.3.2 Practical Issues 304  
9.3.2.1 Unequal Sample Sizes, Missing Data, and Power 304  
9.3.2.2 Multivariate Normality 305  
9.3.2.3 Absence of Outliers 305  
9.3.2.4 Homogeneity of Variance–Covariance Matrices 305  
9.3.2.5 Linearity 306  
9.3.2.6 Absence of Multicollinearity and Singularity 306  
9.4 Fundamental Equations for Discriminant Analysis 306  
9.4.1 Derivation and Test of Discriminant Functions 307  
9.4.2 Classification 309  
9.4.3 Computer Analyses of Small-Sample Example 311  
9.5 Types of Discriminant Analyses 315  
9.5.1 Direct Discriminant Analysis 315  
9.5.2 Sequential Discriminant Analysis 315  
9.5.3 Stepwise (Statistical) Discriminant Analysis 316  
9.6 Some Important Issues 316  
9.6.1 Statistical Inference 316  
9.6.1.1 Criteria for Overall Statistical Significance 317  
9.6.1.2 Stepping Methods 317  
9.6.2 Number of Discriminant Functions 317  
9.6.3 Interpreting Discriminant Functions 318  
9.6.3.1 Discriminant Function Plots 318  
9.6.3.2 Structure Matrix of Loadings 318  
9.6.4 Evaluating Predictor Variables 320  
9.6.5 Effect Size 321  
9.6.6 Design Complexity: Factorial Designs 321  
9.6.7 Use of Classification Procedures 322  
9.6.7.1 Cross-Validation and New Cases 322  
9.6.7.2 Jackknifed Classification 323  
9.6.7.3 Evaluating Improvement in Classification 323  
9.7 Complete Example of Discriminant Analysis 324  
9.7.1 Evaluation of Assumptions 325  
9.7.1.1 Unequal Sample Sizes and Missing Data 325  
9.7.1.2 Multivariate Normality 325  
9.7.1.3 Linearity 325  
9.7.1.4 Outliers 325  
9.7.1.5 Homogeneity of Variance–Covariance Matrices 326  
9.7.1.6 Multicollinearity and Singularity 327  
9.7.2 Direct Discriminant Analysis 327  
9.8 Comparison of Programs 340  
9.8.1 IBM SPSS Package 344  
9.8.2 SAS System 344  
9.8.3 SYSTAT System 345  
10 Logistic Regression  
10.1 General Purpose and Description 346  
10.2 Kinds of Research Questions 348  
10.2.1 Prediction of Group Membership or Outcome 348  
10.2.2 Importance of Predictors 348  
10.2.3 Interactions Among Predictors 349  
10.2.4 Parameter Estimates 349  
10.2.5 Classification of Cases 349  
10.2.6 Significance of Prediction with Covariates 349  
10.2.7 Effect Size 349  
10.3 Limitations to Logistic Regression Analysis 350  
10.3.1 Theoretical Issues 350  
10.3.2 Practical Issues 350  
10.3.2.1 Ratio of Cases to Variables 350  
10.3.2.2 Adequacy of Expected Frequencies and Power 351  
10.3.2.3 Linearity in the Logit 351  
10.3.2.4 Absence of Multicollinearity 351  
10.3.2.5 Absence of Outliers in the Solution 351  
10.3.2.6 Independence of Errors 352  
10.4 Fundamental Equations for Logistic Regression 352  
10.4.1 Testing and Interpreting Coefficients 353  
10.4.2 Goodness of Fit 354  
10.4.3 Comparing Models 355  
10.4.4 Interpretation and Analysis of Residuals 355  
10.4.5 Computer Analyses of Small-Sample Example 356  
10.5 Types of Logistic Regression 360  
10.5.1 Direct Logistic Regression 360  
10.5.2 Sequential Logistic Regression 360  
10.5.3 Statistical (Stepwise) Logistic Regression 362  
10.5.4 Probit and Other Analyses 362  
10.6 Some Important Issues 363  
10.6.1 Statistical Inference 363  
10.6.1.1 Assessing Goodness of Fit of Models 363  
10.6.1.2 Tests of Individual Predictors 365  
10.6.2 Effect Sizes 365  
10.6.2.1 Effect Size for a Model 365  
10.6.2.2 Effect Sizes for Predictors 366  
10.6.3 Interpretation of Coefficients Using Odds 367  
10.6.4 Coding Outcome and Predictor Categories 368  
10.6.5 Number and Type of Outcome Categories 369  
10.6.6 Classification of Cases 372  
10.6.7 Hierarchical and Nonhierarchical Analysis 372
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
14 Structural Equation Modeling by Jodie B. Ullman 528

14.1 General Purpose and Description 528
14.2 Kinds of Research Questions 531
14.2.1 Adequacy of the Model 531
14.2.2 Testing Theory 531
14.2.3 Amount of Variance in the Variables Accounted for by the Factors 532
14.2.4 Reliability of the Indicators 532
14.2.5 Parameter Estimates 532
14.2.6 Intervening Variables 532
14.2.7 Group Differences 532
14.2.8 Longitudinal Differences 532
14.2.9 Multilevel Modeling 533
14.2.10 Latent Class Analysis 533

14.3 Limitations to Structural Equation Modeling 533
14.3.1 Theoretical Issues 533
14.3.2 Practical Issues 534
14.3.2.1 Sample Size and Missing Data 534
14.3.2.2 Multivariate Normality and Outliers 534
14.3.2.3 Linearity 534
14.3.2.4 Absence of Multicollinearity and Singularity 535
14.3.2.5 Residuals 535

14.4 Fundamental Equations for Structural Equations Modeling 535
14.4.1 Covariance Algebra 535
14.4.2 Model Hypotheses 537
14.4.3 Model Specification 538
14.4.4 Model Estimation 540
14.4.5 Model Evaluation 543
14.4.6 Computer Analysis of Small-Sample Example 545

14.5 Some Important Issues 555
14.5.1 Model Identification 555
14.5.2 Estimation Techniques 557
14.5.2.1 Estimation Methods and Sample Size 559
14.5.2.2 Estimation Methods and Nonnormality 559
14.5.2.3 Estimation Methods and Dependence 559
14.5.2.4 Some Recommendations for Choice of Estimation Method 560
14.5.3 Assessing the Fit of the Model 560
14.5.3.1 Comparative Fit Indices 560
14.5.3.2 Absolute Fit Index 562

14.5.3.3 Indices of Proportion of Variance Accounted 562
14.5.3.4 Degree of Parsimony Fit Indices 563
14.5.3.5 Residual-Based Fit Indices 563
14.5.3.6 Choosing Among Fit Indices 564

14.5.4 Model Modification 564
14.5.4.1 Chi-Square Difference Test 564
14.5.4.2 Lagrange Multiplier (LM) Test 565
14.5.4.3 Wald Test 569
14.5.4.4 Some Caveats and Hints on Model Modification 570

14.5.5 Reliability and Proportion of Variance 570
14.5.6 Discrete and Ordinal Data 571
14.5.7 Multiple Group Models 572
14.5.8 Mean and Covariance Structure Models 573

14.6 Complete Examples of Structural Equation Modeling Analysis 574
14.6.1 Confirmatory Factor Analysis of the WISC 574
14.6.1.1 Model Specification for CFA 574
14.6.1.2 Evaluation of Assumptions for CFA 574
14.6.1.3 CFA Model Estimation and Preliminary Evaluation 576
14.6.1.4 Model Modification 583
14.6.2 SEM of Health Data 589
14.6.2.1 SEM Model Specification 589
14.6.2.2 Evaluation of Assumptions for SEM 591
14.6.2.3 SEM Model Estimation and Preliminary Evaluation 593
14.6.2.4 Model Modification 596

14.7 Comparison of Programs 607
14.7.1 EQS 607
14.7.2 LISREL 607
14.7.3 AMOS 612
14.7.4 SAS System 612

15 Multilevel Linear Modeling 613

15.1 General Purpose and Description 613
15.2 Kinds of Research Questions 616
15.2.1 Group Differences in Means 616
15.2.2 Group Differences in Slopes 616
15.2.3 Cross-Level Interactions 616
15.2.4 Meta-Analysis 616
15.2.5 Relative Strength of Predictors at Various Levels 617
15.2.6 Individual and Group Structure 617
15.2.7 Effect Size 617
15.2.8 Path Analysis at Individual and Group Levels 617
15.2.9 Analysis of Longitudinal Data 617
15.2.10 Multilevel Logistic Regression 618
15.2.11 Multiple Response Analysis 618
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 Model with Predictors at Both Levels 633
    15.4.3.2 Level-2 Equations for 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
Appendix B
Research Designs for Complete Examples

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

Appendix C
Statistical Tables 797

C.1 Normal Curve Areas 798
C.2 Critical Values of the t Distribution for α = .05 and .01, Two-Tailed Test 799
C.3 Critical Values of the F Distribution 800
C.4 Critical Values of Chi Square (χ²) 804
C.5 Critical Values for Squares Multiple Correlation (R²) in Forward Stepwise Selection: α = .05 805
C.6 Critical Values for FMAX (S²MAX/S²MIN) Distribution for α = .05 and .01 807

References 808
Index 815
Preface

Some good things seem to go on forever: friendship and updating this book. It is difficult 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 importance has been incorporated in multiple regression, canonical correlation, and logistic regression 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 complex 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 understanding. 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 underlying the technique, which some readers truly enjoy working through (we know because they helpfully 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 issues 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 coffers 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
Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest. Uses for multivariate analysis include: design for capability (also known as capability-based design). This package contains descriptive statistics for multivariate data and distributions derived from the multivariate normal distribution. Distributions are represented in the symbolic form name[param_1,param_2,\ldots][Ellipsis]. This loads the package. Here is a bivariate dataset (courtesy of United States Forest Products Laboratory).