CONTENTS

Preface xxxi

1 Introduction 1

1.1 Multivariate Statistics: Why? 1
1.1.1 The Domain of Multivariate Statistics: Numbers of IVs and DVs 1
1.1.2 Experimental and Nonexperimental Research 2
1.1.3 Computers and Multivariate Statistics 3
1.1.4 Garbage In, Roses Out? 5

1.2 Some Useful Definitions 5
1.2.1 Continuous, Discrete, and Dichotomous Data 5
1.2.2 Samples and Populations 7
1.2.3 Descriptive and Inferential Statistics 7
1.2.4 Orthogonality: Standard and Sequential Analyses 8

1.3 Linear Combinations of Variables 10

1.4 Number and Nature of Variables to Include 11

1.5 Statistical Power 11

1.6 Data Appropriate for Multivariate Statistics 12
1.6.1 The Data Matrix 12
1.6.2 The Correlation Matrix 13
1.6.3 The Variance–Covariance Matrix 14
1.6.4 The Sum-of-Squares and Cross-Products Matrix 14
1.6.5 Residuals 16

1.7 Organization of the Book 16

2 A Guide to Statistical Techniques: Using the Book 17

2.1 Research Questions and Associated Techniques 17
2.1.1 Degree of Relationship Among Variables 17
2.1.1.1 Bivariate *r* 17
2.1.1.2 Multiple *R* 18
2.1.1.3 Sequential *R* 18
2.1.1.4 Canonical *R* 18
2.1.1.5 Multiway Frequency Analysis 19
2.1.1.6 Multilevel Modeling 19

2.1.2 Significance of Group Differences 19
2.1.2.1 One-Way ANOVA and *t* Test 19
2.1.2.2 One-Way ANCOVA 19
2.1.2.3 Factorial ANOVA 20
2.1.2.4 Factorial ANCOVA 20
2.1.2.5 Hotelling's T^2 20
2.1.2.6 One-Way MANOVA 21
2.1.2.7 One-Way MANCOVA 21
2.1.2.8 Factorial MANOVA 22
2.1.2.9 Factorial MANCOVA 22
2.1.2.10 Profile Analysis of Repeated Measures 22
2.1.3 Prediction of Group Membership 23
2.1.3.1 One-Way Discriminant Analysis 23
2.1.3.2 Sequential One-Way Discriminant Analysis 24
2.1.3.3 Multiway Frequency Analysis (Logit) 24
2.1.3.4 Logistic Regression 24
2.1.3.5 Sequential Logistic Regression 24
2.1.3.6 Factorial Discriminant Analysis 25
2.1.3.7 Sequential Factorial Discriminant Analysis 25
2.1.4 Structure 25
2.1.4.1 Principal Components 25
2.1.4.2 Factor Analysis 25
2.1.4.3 Structural Equation Modeling 26
2.1.5 Time Course of Events 26
2.1.5.1 Survival/Failure Analysis 26
2.1.5.2 Time-Series Analysis 26

2.2 Some Further Comparisons 27

2.3 A Decision Tree 28

2.4 Technique Chapters 31

2.5 Preliminary Check of the Data 32

3 Review of Univariate and Bivariate Statistics 33

3.1 Hypothesis Testing 33
3.1.1 One-Sample *z* Test as Prototype 33
3.1.2 Power 36
3.1.3 Extensions of the Model 37
3.1.4 Controversy Surrounding Significance Testing 37

3.2 Analysis of Variance 37
3.2.1 One-Way Between-Subjects ANOVA 39
3.2.2 Factorial Between-Subjects ANOVA 42
3.2.3 Within-Subjects ANOVA 43
3.2.4 Mixed Between-Within-Subjects ANOVA 46

3.2.5 Design Complexity 47
3.2.5.1 Nesting 47
3.2.5.2 Latin-Square Designs 47
3.2.5.3 Unequal *n* and Nonorthogonality 48
3.2.5.4 Fixed and Random Effects 49
3.2.6 Specific Comparisons 49
3.2.6.1 Weighting Coefficients for Comparisons 50
3.2.6.2 Orthogonality of Weighting Coefficients 50
3.2.6.3 Obtained *F* for Comparisons 51
3.2.6.4 Critical *F* for Planned Comparisons 52
3.2.6.5 Critical *F* for Post Hoc Comparisons 52

3.3 Parameter Estimation 53

3.4 Effect Size 54

3.5 Bivariate Statistics: Correlation and Regression 55
3.5.1 Correlation 56
3.5.2 Regression 57

3.6 Chi-Square Analysis 58

4 Cleaning Up Your Act: Screening Data Prior to Analysis 60

4.1 Important Issues in Data Screening 61
4.1.1 Accuracy of Data File 61
4.1.2 Honest Correlations 61
4.1.2.1 Inflated Correlation 61
4.1.2.2 Deflated Correlation 61
4.1.3 Missing Data 62
4.1.3.1 Deleting Cases or Variables 63
4.1.3.2 Estimating Missing Data 66
4.1.3.3 Using a Missing Data Correlation Matrix 70
4.1.3.4 Treating Missing Data as Data 71
4.1.3.5 Repeating Analyses With and Without Missing Data 71
4.1.3.6 Choosing Among Methods for Dealing With Missing Data 71
4.1.4 Outliers 72
4.1.4.1 Detecting Univariate and Multivariate Outliers 73
4.1.4.2 Describing Outliers 76
4.1.4.3 Reducing the Influence of Outliers 77
4.1.4.4 Outliers in a Solution 77
4.1.5 Normality, Linearity, and Homoscedasticity 78
4.1.5.1 Normality 79
4.1.5.2 Linearity 83
4.1.5.3 Homoscedasticity, Homogeneity of Variance, and Homogeneity of Variance–Covariance Matrices 85

4.1.6 Common Data Transformations 86
4.1.7 Multicollinearity and Singularity 88
4.1.8 A Checklist and Some Practical Recommendations 91

4.2 Complete Examples of Data Screening 92
4.2.1 Screening Ungrouped Data 92
4.2.1.1 Accuracy of Input, Missing Data, Distributions, and Univariate Outliers 93
4.2.1.2 Linearity and Homoscedasticity 97
4.2.1.3 Transformation 98
4.2.1.4 Detecting Multivariate Outliers 99
4.2.1.5 Variables Causing Cases to Be Outliers 100
4.2.1.6 Multicollinearity 104
4.2.2 Screening Grouped Data 105
4.2.2.1 Accuracy of Input, Missing Data, Distributions, Homogeneity of Variance, and Univariate Outliers 105
4.2.2.2 Linearity 110
4.2.2.3 Multivariate Outliers 111
4.2.2.4 Variables Causing Cases to Be Outliers 113
4.2.2.5 Multicollinearity 114

5 Multiple Regression 117

5.1 General Purpose and Description 117

5.2 Kinds of Research Questions 119
5.2.1 Degree of Relationship 119
5.2.2 Importance of IVs 120
5.2.3 Adding IVs 120
5.2.4 Changing IVs 120
5.2.5 Contingencies Among IVs 120
5.2.6 Comparing Sets of IVs 121
5.2.7 Predicting DV Scores for Members of a New Sample 121
5.2.8 Parameter Estimates 121

5.3 Limitations to Regression Analyses 122
5.3.1 Theoretical Issues 122
5.3.2 Practical Issues 123
5.3.2.1 Ratio of Cases to IVs 123
5.3.2.2 Absence of Outliers Among the IVs and on the DV 124
5.3.2.3 Absence of Multicollinearity and Singularity 125
5.3.2.4 Normality, Linearity, and Homoscedasticity of Residuals 125
5.3.2.5 Independence of Errors 128
5.3.2.6 Absence of Outliers in the Solution 128

5.4 Fundamental Equations for Multiple Regression 129
5.4.1 General Linear Equations 129
5.4.2 Matrix Equations 131
5.4.3 Computer Analyses of Small-Sample Example 133

5.5 Major Types of Multiple Regression 136
5.5.1 Standard Multiple Regression 136
5.5.2 Sequential Multiple Regression 137
5.5.3 Statistical (Stepwise) Regression 138
5.5.4 Choosing Among Regression Strategies 143

5.6 Some Important Issues 144
5.6.1 Importance of IVs 144
5.6.1.1 Standard Multiple Regression 144
5.6.1.2 Sequential or Statistical Regression 145
5.6.1.3 Commonality Analysis 146
5.6.2 Statistical Inference 149
5.6.2.1 Test for Multiple R 149
5.6.2.2 Test of Regression Components 150
5.6.2.3 Test of Added Subset of IVs 151
5.6.2.4 Confidence Limits Around B and Multiple R^2 151
5.6.2.5 Comparing Two Sets of Predictors 153
5.6.3 Adjustment of R^2 154
5.6.4 Suppressor Variables 155
5.6.5 Regression Approach to ANOVA 156
5.6.6 Centering When Interactions and Powers of IVs Are Included 158
5.6.7 Mediation in Causal Sequence 160

5.7 Complete Examples of Regression Analysis 161
5.7.1 Evaluation of Assumptions 162
5.7.1.1 Ratio of Cases to IVs 162
5.7.1.2 Normality, Linearity, Homoscedasticity, and Independence of Residuals 162
5.7.1.3 Outliers 167
5.7.1.4 Multicollinearity and Singularity 168
5.7.2 Standard Multiple Regression 169
5.7.3 Sequential Regression 175
5.7.4 Example of Standard Multiple Regression With Missing Values Multiply Imputed 181

5.8 Comparison of Programs 190
5.8.1 IBM SPSS Package 190
5.8.2 SAS System 195
5.8.3 SYSTAT System 196

6 Analysis of Covariance 197

6.1 General Purpose and Description 197

6.2 Kinds of Research Questions 200
- 6.2.1 Main Effects of IVs 200
- 6.2.2 Interactions Among IVs 200
- 6.2.3 Specific Comparisons and Trend Analysis 201
- 6.2.4 Effects of Covariates 201
- 6.2.5 Effect Size 201
- 6.2.6 Parameter Estimates 201

6.3 Limitations to Analysis of Covariance 202
- 6.3.1 Theoretical Issues 202
- 6.3.2 Practical Issues 203
  - 6.3.2.1 Unequal Sample Sizes, Missing Data, and Ratio of Cases to IVs 203
  - 6.3.2.2 Absence of Outliers 203
  - 6.3.2.3 Absence of Multicollinearity and Singularity 203
  - 6.3.2.4 Normality of Sampling Distributions 204
  - 6.3.2.5 Homogeneity of Variance 204
  - 6.3.2.6 Linearity 204
  - 6.3.2.7 Homogeneity of Regression 204
  - 6.3.2.8 Reliability of Covariates 205

6.4 Fundamental Equations for Analysis of Covariance 205
- 6.4.1 Sums of Squares and Cross-Products 206
- 6.4.2 Significance Test and Effect Size 210
- 6.4.3 Computer Analyses of Small-Sample Example 211

6.5 Some Important Issues 213
- 6.5.1 Choosing Covariates 213
- 6.5.2 Evaluation of Covariates 214
- 6.5.3 Test for Homogeneity of Regression 215
- 6.5.4 Design Complexity 215
  - 6.5.4.1 Within-Subjects and Mixed Within-Between Designs 216
  - 6.5.4.2 Unequal Sample Sizes 219
  - 6.5.4.3 Specific Comparisons and Trend Analysis 220
  - 6.5.4.4 Effect Size 223
- 6.5.5 Alternatives to ANCOVA 223

6.6 Complete Example of Analysis of Covariance 225
- 6.6.1 Evaluation of Assumptions 225
  - 6.6.1.1 Unequal n and Missing Data 226
  - 6.6.1.2 Normality 226
  - 6.6.1.3 Linearity 226
  - 6.6.1.4 Outliers 226
  - 6.6.1.5 Multicollinearity and Singularity 230
  - 6.6.1.6 Homogeneity of Variance 230

6.6.1.7 Homogeneity of Regression 232
6.6.1.8 Reliability of Covariates 232
6.6.2 Analysis of Covariance 232
6.6.2.1 Main Analysis 232
6.6.2.2 Evaluation of Covariates 234
6.6.2.3 Homogeneity of Regression Run 239

6.7 Comparison of Programs 242
6.7.1 SPSS Package 242
6.7.2 SAS System 242
6.7.3 SYSTAT System 242

7 Multivariate Analysis of Variance and Covariance 245

7.1 General Purpose and Description 245

7.2 Kinds of Research Questions 248
7.2.1 Main Effects of IVs 249
7.2.2 Interactions Among IVs 249
7.2.3 Importance of DVs 249
7.2.4 Parameter Estimates 250
7.2.5 Specific Comparisons and Trend Analysis 250
7.2.6 Effect Size 250
7.2.7 Effects of Covariates 250
7.2.8 Repeated-Measures Analysis of Variance 251

7.3 Limitations to Multivariate Analysis of Variance and Covariance 251
7.3.1 Theoretical Issues 251
7.3.2 Practical Issues 252
7.3.2.1 Unequal Sample Sizes, Missing Data, and Power 252
7.3.2.2 Multivariate Normality 252
7.3.2.3 Absence of Outliers 253
7.3.2.4 Homogeneity of Variance–Covariance Matrices 253
7.3.2.5 Linearity 254
7.3.2.6 Homogeneity of Regression 254
7.3.2.7 Reliability of Covariates 255
7.3.2.8 Absence of Multicollinearity and Singularity 255

7.4 Fundamental Equations for Multivariate Analysis of Variance and Covariance 255
7.4.1 Multivariate Analysis of Variance 255
7.4.2 Computer Analyses of Small-Sample Example 263
7.4.3 Multivariate Analysis of Covariance 266

7.5 Some Important Issues 270
7.5.1 MANOVA versus ANOVAs 270
7.5.2 Criteria for Statistical Inference 270

7.5.3 Assessing DVs 271
7.5.3.1 Univariate *F* 272
7.5.3.2 Roy–Bargmann Stepdown Analysis 273
7.5.3.3 Using Discriminant Analysis 274
7.5.3.4 Choosing Among Strategies for Assessing DVs 275
7.5.4 Specific Comparisons and Trend Analysis 275
7.5.5 Design Complexity 276
7.5.5.1 Within-Subjects and Between-Within Designs 276
7.5.5.2 Unequal Sample Sizes 278

7.6 Complete Examples of Multivariate Analysis of Variance and Covariance 279
7.6.1 Evaluation of Assumptions 279
7.6.1.1 Unequal Sample Sizes and Missing Data 279
7.6.1.2 Multivariate Normality 281
7.6.1.3 Linearity 281
7.6.1.4 Outliers 282
7.6.1.5 Homogeneity of Variance–Covariance Matrices 282
7.6.1.6 Homogeneity of Regression 283
7.6.1.7 Reliability of Covariates 286
7.6.1.8 Multicollinearity and Singularity 287
7.6.2 Multivariate Analysis of Variance 287
7.6.3 Multivariate Analysis of Covariance 298
7.6.3.1 Assessing Covariates 298
7.6.3.2 Assessing DVs 300

7.7 Comparison of Programs 310
7.7.1 IBM SPSS Package 310
7.7.2 SAS System 313
7.7.3 SYSTAT System 313

8 Profile Analysis: The Multivariate Approach to Repeated Measures 314

8.1 General Purpose and Description 314

8.2 Kinds of Research Questions 315
8.2.1 Parallelism of Profiles 315
8.2.2 Overall Difference Among Groups 316
8.2.3 Flatness of Profiles 316
8.2.4 Contrasts Following Profile Analysis 316
8.2.5 Parameter Estimates 316
8.2.6 Effect Size 316

8.3 Limitations to Profile Analysis 317
8.3.1 Theoretical Issues 317
8.3.2 Practical Issues 317
8.3.2.1 Sample Size, Missing Data, and Power 317
8.3.2.2 Multivariate Normality 318

8.3.2.3 Absence of Outliers 318
8.3.2.4 Homogeneity of Variance–Covariance Matrices 318
8.3.2.5 Linearity 318
8.3.2.6 Absence of Multicollinearity and Singularity 319

8.4 Fundamental Equations for Profile Analysis 319
8.4.1 Differences in Levels 320
8.4.2 Parallelism 321
8.4.3 Flatness 324
8.4.4 Computer Analyses of Small-Sample Example 325

8.5 Some Important Issues 331
8.5.1 Univariate Versus Multivariate Approach to Repeated Measures 331
8.5.2 Contrasts in Profile Analysis 333
8.5.2.1 Parallelism and Flatness Significant, Levels Not Significant (Simple-Effects Analysis) 335
8.5.2.2 Parallelism and Levels Significant, Flatness Not Significant (Simple-Effects Analysis) 337
8.5.2.3 Parallelism, Levels, and Flatness Significant (Interaction Contrasts) 341
8.5.2.4 Only Parallelism Significant 341
8.5.3 Doubly Multivariate Designs 343
8.5.4 Classifying Profiles 347
8.5.5 Imputation of Missing Values 347

8.6 Complete Examples of Profile Analysis 348
8.6.1 Profile Analysis of Subscales of the WISC 348
8.6.1.1 Evaluation of Assumptions 349
8.6.1.2 Profile Analysis 353
8.6.2 Doubly Multivariate Analysis of Reaction Time 362
8.6.2.1 Evaluation of Assumptions 362
8.6.2.2 Doubly Multivariate Analysis of Slope and Intercept 365

8.7 Comparison of Programs 373
8.7.1 IBM SPSS Package 374
8.7.2 SAS System 374
8.7.3 SYSTAT System 376

9 Discriminant Analysis 377

9.1 General Purpose and Description 377

9.2 Kinds of Research Questions 380
9.2.1 Significance of Prediction 380
9.2.2 Number of Significant Discriminant Functions 380
9.2.3 Dimensions of Discrimination 381
9.2.4 Classification Functions 381
9.2.5 Adequacy of Classification 381

9.2.6 Effect Size 381
9.2.7 Importance of Predictor Variables 382
9.2.8 Significance of Prediction With Covariates 382
9.2.9 Estimation of Group Means 382

9.3 Limitations to Discriminant Analysis 383
9.3.1 Theoretical Issues 383
9.3.2 Practical Issues 383
9.3.2.1 Unequal Sample Sizes, Missing Data, and Power 383
9.3.2.2 Multivariate Normality 384
9.3.2.3 Absence of Outliers 384
9.3.2.4 Homogeneity of Variance–Covariance Matrices 384
9.3.2.5 Linearity 385
9.3.2.6 Absence of Multicollinearity and Singularity 385

9.4 Fundamental Equations for Discriminant Analysis 386
9.4.1 Derivation and Test of Discriminant Functions 386
9.4.2 Classification 389
9.4.3 Computer Analyses of Small-Sample Example 391

9.5 Types of Discriminant Analyses 397
9.5.1 Direct Discriminant Analysis 397
9.5.2 Sequential Discriminant Analysis 398
9.5.3 Stepwise (Statistical) Discriminant Analysis 398

9.6 Some Important Issues 399
9.6.1 Statistical Inference 399
9.6.1.1 Criteria for Overall Statistical Significance 399
9.6.1.2 Stepping Methods 399
9.6.2 Number of Discriminant Functions 400
9.6.3 Interpreting Discriminant Functions 400
9.6.3.1 Discriminant Function Plots 400
9.6.3.2 Structure Matrix of Loadings 402
9.6.4 Evaluating Predictor Variables 403
9.6.5 Effect Size 404
9.6.6 Design Complexity: Factorial Designs 405
9.6.7 Use of Classification Procedures 406
9.6.7.1 Cross-Validation and New Cases 407
9.6.7.2 Jackknifed Classification 407
9.6.7.3 Evaluating Improvement in Classification 407

9.7 Complete Example of Discriminant Analysis 409
9.7.1 Evaluation of Assumptions 409
9.7.1.1 Unequal Sample Sizes and Missing Data 409
9.7.1.2 Multivariate Normality 410
9.7.1.3 Linearity 410
9.7.1.4 Outliers 410
9.7.1.5 Homogeneity of Variance–Covariance Matrices 413
9.7.1.6 Multicollinearity and Singularity 413
9.7.2 Direct Discriminant Analysis 414

9.8 Comparison of Programs 432
9.8.1 IBM SPSS Package 432
9.8.2 SAS System 432
9.8.3 SYSTAT System 438

10 Logistic Regression 439

10.1 General Purpose and Description 439

10.2 Kinds of Research Questions 441
10.2.1 Prediction of Group Membership or Outcome 441
10.2.2 Importance of Predictors 441
10.2.3 Interactions among Predictors 442
10.2.4 Parameter Estimates 442
10.2.5 Classification of Cases 442
10.2.6 Significance of Prediction with Covariates 442
10.2.7 Effect Size 443

10.3 Limitations to Logistic Regression Analysis 443
10.3.1 Theoretical Issues 443
10.3.2 Practical Issues 444
10.3.2.1 Ratio of Cases to Variables 444
10.3.2.2 Adequacy of Expected Frequencies and Power 444
10.3.2.3 Linearity in the Logit 445
10.3.2.4 Absence of Multicollinearity 445
10.3.2.5 Absence of Outliers in the Solution 445
10.3.2.6 Independence of Errors 445

10.4 Fundamental Equations for Logistic Regression 446
10.4.1 Testing and Interpreting Coefficients 447
10.4.2 Goodness of Fit 448
10.4.3 Comparing Models 450
10.4.4 Interpretation and Analysis of Residuals 450
10.4.5 Computer Analyses of Small-Sample Example 451

10.5 Types of Logistic Regression 455
10.5.1 Direct Logistic Regression 456
10.5.2 Sequential Logistic Regression 456
10.5.3 Statistical (Stepwise) Logistic Regression 456
10.5.4 Probit and Other Analyses 458

10.6 Some Important Issues 459
10.6.1 Statistical Inference 459
10.6.1.1 Assessing Goodness of Fit of Models 459
10.6.1.2 Tests of Individual Variables 461
10.6.2 Effect Size for a Model 462
10.6.3 Interpretation of Coefficients Using Odds 463
10.6.4 Coding Outcome and Predictor Categories 465

10.6.5 Number and Type of Outcome Categories 466
10.6.6 Classification of Cases 469
10.6.7 Hierarchical and Nonhierarchical Analysis 470
10.6.8 Importance of Predictors 472
10.6.9 Logistic Regression for Matched Groups 472

10.7 Complete Examples of Logistic Regression 472
10.7.1 Evaluation of Limitations 473
10.7.1.1 Ratio of Cases to Variables and Missing Data 473
10.7.1.2 Multicollinearity 476
10.7.1.3 Outliers in the Solution 477
10.7.2 Direct Logistic Regression with Two-Category Outcome and Continuous Predictors 477
10.7.2.1 Limitation: Linearity in the Logit 477
10.7.2.2 Direct Logistic Regression With Two-Category Outcome 477
10.7.3 Sequential Logistic Regression with Three Categories of Outcome 484
10.7.3.1 Limitations of Multinomial Logistic Regression 484
10.7.3.2 Sequential Multinomial Logistic Regression 484

10.8 Comparison of Programs 502
10.8.1 IBM SPSS Package 502
10.8.2 SAS System 508
10.8.3 SYSTAT System 509

11 Survival/Failure Analysis 510

11.1 General Purpose and Description 510

11.2 Kinds of Research Questions 512
11.2.1 Proportions Surviving at Various Times 512
11.2.2 Group Differences in Survival 512
11.2.3 Survival Time With Covariates 512
11.2.3.1 Treatment Effects 512
11.2.3.2 Importance of Covariates 512
11.2.3.3 Parameter Estimates 513
11.2.3.4 Contingencies Among Covariates 513
11.2.3.5 Effect Size and Power 513

11.3 Limitations to Survival Analysis 513
11.3.1 Theoretical Issues 513
11.3.2 Practical Issues 513
11.3.2.1 Sample Size and Missing Data 514
11.3.2.2 Normality of Sampling Distributions, Linearity, and Homoscedasticity 514
11.3.2.3 Absence of Outliers 514
11.3.2.4 Differences Between Withdrawn and Remaining Cases 514

11.3.2.5 Change in Survival Conditions over Time 514
11.3.2.6 Proportionality of Hazards 515
11.3.2.7 Absence of Multicollinearity 515

11.4 Fundamental Equations for Survival Analysis 515
11.4.1 Life Tables 516
11.4.2 Standard Error of Cumulative Proportion Surviving 518
11.4.3 Hazard and Density Functions 518
11.4.4 Plot of Life Tables 519
11.4.5 Test for Group Differences 520
11.4.6 Computer Analyses of Small-Sample Example 522

11.5 Types of Survival Analyses 528
11.5.1 Actuarial and Product-Limit Life Tables and Survivor Functions 528
11.5.2 Prediction of Group Survival Times From Covariates 529
11.5.2.1 Direct, Sequential, and Statistical Analysis 531
11.5.2.2 Cox Proportional-Hazards Model 531
11.5.2.3 Accelerated Failure-Time Models 532
11.5.2.4 Choosing a Method 539

11.6 Some Important Issues 539
11.6.1 Proportionality of Hazards 539
11.6.2 Censored Data 541
11.6.2.1 Right-Censored Data 541
11.6.2.2 Other Forms of Censoring 541
11.6.3 Effect Size and Power 542
11.6.4 Statistical Criteria 543
11.6.4.1 Test Statistics for Group Differences in Survival Functions 543
11.6.4.2 Test Statistics for Prediction From Covariates 544
11.6.5 Predicting Survival Rate 544
11.6.5.1 Regression Coefficients (Parameter Estimates) 544
11.6.5.2 Hazard Ratios 544
11.6.5.3 Expected Survival Rates 545

11.7 Complete Example of Survival Analysis 545
11.7.1 Evaluation of Assumptions 547
11.7.1.1 Accuracy of Input, Adequacy of Sample Size, Missing Data, and Distributions 547
11.7.1.2 Outliers 549
11.7.1.3 Differences Between Withdrawn and Remaining Cases 553
11.7.1.4 Change in Survival Experience over Time 553
11.7.1.5 Proportionality of Hazards 553
11.7.1.6 Multicollinearity 555
11.7.2 Cox Regression Survival Analysis 555
11.7.2.1 Effect of Drug Treatment 556
11.7.2.2 Evaluation of Other Covariates 556

11.8 Comparison of Programs 563
11.8.1 SAS System 563
11.8.2 IBM SPSS Package 569
11.8.3 SYSTAT System 570

12 Canonical Correlation 571

12.1 General Purpose and Description 571

12.2 Kinds of Research Questions 573
12.2.1 Number of Canonical Variate Pairs 573
12.2.2 Interpretation of Canonical Variates 573
12.2.3 Importance of Canonical Variates 573
12.2.4 Canonical Variate Scores 574

12.3 Limitations 574
12.3.1 Theoretical Limitations 574
12.3.2 Practical Issues 575
12.3.2.1 Ratio of Cases to IVs 575
12.3.2.2 Normality, Linearity, and Homoscedasticity 575
12.3.2.3 Missing Data 576
12.3.2.4 Absence of Outliers 576
12.3.2.5 Absence of Multicollinearity and Singularity 576

12.4 Fundamental Equations for Canonical Correlation 576
12.4.1 Eigenvalues and Eigenvectors 578
12.4.2 Matrix Equations 580
12.4.3 Proportions of Variance Extracted 584
12.4.4 Computer Analyses of Small-Sample Example 585

12.5 Some Important Issues 591
12.5.1 Importance of Canonical Variates 591
12.5.2 Interpretation of Canonical Variates 592

12.6 Complete Example of Canonical Correlation 592
12.6.1 Evaluation of Assumptions 593
12.6.1.1 Missing Data 593
12.6.1.2 Normality, Linearity, and Homoscedasticity 593
12.6.1.3 Outliers 595
12.6.1.4 Multicollinearity and Singularity 595
12.6.2 Canonical Correlation 595

12.7 Comparison of Programs 609
12.7.1 SAS System 609
12.7.2 IBM SPSS Package 609
12.7.3 SYSTAT System 611

13 Principal Components and Factor Analysis 612

13.1 General Purpose and Description 612

13.2 Kinds of Research Questions 615
13.2.1 Number of Factors 615
13.2.2 Nature of Factors 616
13.2.3 Importance of Solutions and Factors 616
13.2.4 Testing Theory in FA 616
13.2.5 Estimating Scores on Factors 616

13.3 Limitations 616
13.3.1 Theoretical Issues 616
13.3.2 Practical Issues 617
13.3.2.1 Sample Size and Missing Data 618
13.3.2.2 Normality 618
13.3.2.3 Linearity 618
13.3.2.4 Absence of Outliers Among Cases 619
13.3.2.5 Absence of Multicollinearity and Singularity 619
13.3.2.6 Factorability of R 619
13.3.2.7 Absence of Outliers Among Variables 620

13.4 Fundamental Equations for Factor Analysis 620
13.4.1 Extraction 622
13.4.2 Orthogonal Rotation 625
13.4.3 Communalities, Variance, and Covariance 626
13.4.4 Factor Scores 627
13.4.5 Oblique Rotation 630
13.4.6 Computer Analyses of Small-Sample Example 632

13.5 Major Types of Factor Analyses 637
13.5.1 Factor Extraction Techniques 637
13.5.1.1 PCA Versus FA 639
13.5.1.2 Principal Components 640
13.5.1.3 Principal Factors 640
13.5.1.4 Image Factor Extraction 641
13.5.1.5 Maximum Likelihood Factor Extraction 641
13.5.1.6 Unweighted Least Squares Factoring 641
13.5.1.7 Generalized (Weighted) Least Squares Factoring 641
13.5.1.8 Alpha Factoring 642
13.5.2 Rotation 642
13.5.2.1 Orthogonal Rotation 642
13.5.2.2 Oblique Rotation 644
13.5.2.3 Geometric Interpretation 645
13.5.3 Some Practical Recommendations 647

13.6 Some Important Issues 647
13.6.1 Estimates of Communalities 648
13.6.2 Adequacy of Extraction and Number of Factors 648
13.6.3 Adequacy of Rotation and Simple Structure 651
13.6.4 Importance and Internal Consistency of Factors 652
13.6.5 Interpretation of Factors 654
13.6.6 Factor Scores 655
13.6.7 Comparisons Among Solutions and Groups 656

13.7 Complete Example of FA 656
13.7.1 Evaluation of Limitations 657
13.7.1.1 Sample Size and Missing Data 657
13.7.1.2 Normality 657
13.7.1.3 Linearity 657
13.7.1.4 Outliers 658
13.7.1.5 Multicollinearity and Singularity 661
13.7.1.6 Factorability of R 661
13.7.1.7 Outliers Among Variables 661
13.7.2 Principal Factors Extraction With Varimax Rotation 661

13.8 Comparison of Programs 676
13.8.1 IBM SPSS Package 676
13.8.2 SAS System 676
13.8.3 SYSTAT System 680

14 Structural Equation Modeling 681
by Jodie B. Ullman

14.1 General Purpose and Description 681

14.2 Kinds of Research Questions 685
14.2.1 Adequacy of the Model 685
14.2.2 Testing Theory 685
14.2.3 Amount of Variance in the Variables Accounted for by the Factors 685
14.2.4 Reliability of the Indicators 685
14.2.5 Parameter Estimates 685
14.2.6 Intervening Variables 686
14.2.7 Group Differences 686
14.2.8 Longitudinal Differences 686
14.2.9 Multilevel Modeling 687
14.2.10 Latent Class Analysis 687

14.3 Limitations to Structural Equation Modeling 687
14.3.1 Theoretical Issues 687
14.3.2 Practical Issues 688
14.3.2.1 Sample Size and Missing Data 688
14.3.2.2 Multivariate Normality and Outliers 688

14.3.2.3 Linearity 689
14.3.2.4 Absence of Multicollinearity and Singularity 689
14.3.2.5 Residuals 689

14.4 Fundamental Equations for Structural Equations Modeling 689
14.4.1 Covariance Algebra 689
14.4.2 Model Hypotheses 691
14.4.3 Model Specification 693
14.4.4 Model Estimation 695
14.4.5 Model Evaluation 699
14.4.6 Computer Analysis of Small-Sample Example 701

14.5 Some Important Issues 714
14.5.1 Model Identification 714
14.5.2 Estimation Techniques 717
14.5.2.1 Estimation Methods and Sample Size 719
14.5.2.2 Estimation Methods and Nonnormality 719
14.5.2.3 Estimation Methods and Dependence 720
14.5.2.4 Some Recommendations for Choice of Estimation Method 720
14.5.3 Assessing the Fit of the Model 720
14.5.3.1 Comparative Fit Indices 721
14.5.3.2 Absolute Fit Index 723
14.5.3.3 Indices of Proportion of Variance Accounted 723
14.5.3.4 Degree of Parsimony Fit Indices 724
14.5.3.5 Residual-Based Fit Indices 725
14.5.3.6 Choosing Among Fit Indices 725
14.5.4 Model Modification 726
14.5.4.1 Chi-Square Difference Test 726
14.5.4.2 Lagrange Multiplier (LM) Test 726
14.5.4.3 Wald Test 728
14.5.4.4 Some Caveats and Hints on Model Modification 733
14.5.5 Reliability and Proportion of Variance 733
14.5.6 Discrete and Ordinal Data 734
14.5.7 Multiple Group Models 735
14.5.8 Mean and Covariance Structure Models 736

14.6 Complete Examples of Structural Equation Modeling Analysis 737
14.6.1 Confirmatory Factor Analysis of the WISC 737
14.6.1.1 Model Specification for CFA 737
14.6.1.2 Evaluation of Assumptions for CFA 738
14.6.1.3 CFA Model Estimation and Preliminary Evaluation 739
14.6.1.4 Model Modification 748
14.6.2 SEM of Health Data 755
14.6.2.1 SEM Model Specification 755
14.6.2.2 Evaluation of Assumptions for SEM 756
14.6.2.3 SEM Model Estimation and Preliminary Evaluation 760
14.6.2.4 Model Modification 764

14.7 Comparison of Programs 778
14.7.1 EQS 778
14.7.2 LISREL 778
14.7.3 AMOS 785
14.7.4 SAS System 785

15 Multilevel Linear Modeling 786

15.1 General Purpose and Description 786

15.2 Kinds of Research Questions 789
15.2.1 Group Differences in Means 789
15.2.2 Group Differences in Slopes 789
15.2.3 Cross-Level Interactions 789
15.2.4 Meta-Analysis 790
15.2.5 Relative Strength of Predictors at Various Levels 790
15.2.6 Individual and Group Structure 790
15.2.7 Effect Size 790
15.2.8 Path Analysis at Individual and Group Levels 790
15.2.9 Analysis of Longitudinal Data 791
15.2.10 Multilevel Logistic Regression 791
15.2.11 Multiple Response Analysis 791

15.3 Limitations to Multilevel Linear Modeling 791
15.3.1 Theoretical Issues 791
15.3.2 Practical Issues 792
15.3.2.1 Sample Size, Unequal-*n*, and Missing Data 792
15.3.2.2 Independence of Errors 793
15.3.2.3 Absence of Multicollinearity and Singularity 794

15.4 Fundamental Equations 794
15.4.1 Intercepts-Only Model 797
15.4.1.1 The Intercepts-Only Model: Level-1 Equation 798
15.4.1.2 The Intercepts-Only Model: Level-2 Equation 798
15.4.1.3 Computer Analyses of Intercepts-Only Model 799
15.4.2 Model With a First-Level Predictor 802
15.4.2.1 Level-1 Equation for a Model With a Level-1 Predictor 803
15.4.2.2 Level-2 Equations for a Model With a Level-1 Predictor 804
15.4.2.3 Computer Analysis of a Model With a Level-1 Predictor 806
15.4.3 Model With Predictors at First and Second Levels 811
15.4.3.1 Level-1 Equation for Model With Predictors at Both Levels 811
15.4.3.2 Level-2 Equations for Model With Predictors at Both Levels 811
15.4.3.3 Computer Analyses of Model With Predictors at First and Second Levels 812

15.5 Types of MLM 818
15.5.1 Repeated Measures 818
15.5.2 Higher-Order MLM 823
15.5.3 Latent Variables 823
15.5.4 Nonnormal Outcome Variables 824
15.5.5 Multiple Response Models 825

15.6 Some Important Issues 826
15.6.1 Intraclass Correlation 826
15.6.2 Centering Predictors and Changes in Their Interpretations 827
15.6.3 Interactions 830
15.6.4 Random and Fixed Intercepts and Slopes 830
15.6.5 Statistical Inference 834
15.6.5.1 Assessing Models 834
15.6.5.2 Tests of Individual Effects 835
15.6.6 Effect Size 836
15.6.7 Estimation Techniques and Convergence Problems 837
15.6.8 Exploratory Model Building 838

15.7 Complete Example of MLM 839
15.7.1 Evaluation of Assumptions 839
15.7.1.1 Sample Sizes, Missing Data, and Distributions 839
15.7.1.2 Outliers 842
15.7.1.3 Multicollinearity and Singularity 843
15.7.1.4 Independence of Errors: Intraclass Correlations 843
15.7.2 Multilevel Modeling 844

15.8 Comparison of Programs 856
15.8.1 SAS System 856
15.8.2 IBM SPSS Package 860
15.8.3 HLM Program 860
15.8.4 MLwiN Program 861
15.8.5 SYSTAT System 861

16 Multiway Frequency Analysis 862

16.1 General Purpose and Description 862

16.2 Kinds of Research Questions 863
16.2.1 Associations Among Variables 863
16.2.2 Effect on a Dependent Variable 864
16.2.3 Parameter Estimates 864
16.2.4 Importance of Effects 864
16.2.5 Effect Size 864
16.2.6 Specific Comparisons and Trend Analysis 865

16.3 Limitations to Multiway Frequency Analysis 865
16.3.1 Theoretical Issues 865
16.3.2 Practical Issues 865
16.3.2.1 Independence 865
16.3.2.2 Ratio of Cases to Variables 866
16.3.2.3 Adequacy of Expected Frequencies 866
16.3.2.4 Absence of Outliers in the Solution 867

16.4 Fundamental Equations for Multiway Frequency Analysis 867
16.4.1 Screening for Effects 868
16.4.1.1 Total Effect 869
16.4.1.2 First-Order Effects 870
16.4.1.3 Second-Order Effects 871
16.4.1.4 Third-Order Effect 875
16.4.2 Modeling 875
16.4.3 Evaluation and Interpretation 878
16.4.3.1 Residuals 878
16.4.3.2 Parameter Estimates 879
16.4.4 Computer Analyses of Small-Sample Example 883

16.5 Some Important Issues 890
16.5.1 Hierarchical and Nonhierarchical Models 890
16.5.2 Statistical Criteria 890
16.5.2.1 Tests of Models 891
16.5.2.2 Tests of Individual Effects 891
16.5.3 Strategies for Choosing a Model 891
16.5.3.1 IBM SPSS HILOGLINEAR (Hierarchical) 892
16.5.3.2 IBM SPSS GENLOG (General Log-Linear) 892
16.5.3.3 SAS CATMOD and IBM SPSS LOGLINEAR (General Log-Linear) 893

16.6 Complete Example of Multiway Frequency Analysis 893
16.6.1 Evaluation of Assumptions: Adequacy of Expected Frequencies 893
16.6.2 Hierarchical Log-Linear Analysis 893
16.6.2.1 Preliminary Model Screening 893
16.6.2.2 Stepwise Model Selection 895
16.6.2.3 Adequacy of Fit 897
16.6.2.4 Interpretation of the Selected Model 901

16.7 Comparison of Programs 910
16.7.1 IBM SPSS Package 913
16.7.2 SAS System 914
16.7.3 SYSTAT System 914

17 An Overview of the General Linear Model 915

17.1 Linearity and the General Linear Model 915

17.2 Bivariate to Multivariate Statistics and Overview of Techniques 915

17.2.1 Bivariate Form 915
17.2.2 Simple Multivariate Form 916
17.2.3 Full Multivariate Form 919

17.3 Alternative Research Strategies 922

18 Time-Series Analysis (available online at www.pearsonhighered.com/tabachnick) 18-1

18.1 General Purpose and Description 18-1

18.2 Kinds of Research Questions 18-3

18.2.1 Pattern of Autocorrelation 18-5
18.2.2 Seasonal Cycles and Trends 18-5
18.2.3 Forecasting 18-5
18.2.4 Effect of an Intervention 18-5
18.2.5 Comparing Time Series 18-5
18.2.6 Time Series With Covariates 18-6
18.2.7 Effect Size and Power 18-6

18.3 Assumptions of Time-Series Analysis 18-6

18.3.1 Theoretical Issues 18-6
18.3.2 Practical Issues 18-6
18.3.2.1 Normality of Distributions of Residuals 18-6
18.3.2.2 Homogeneity of Variance and Zero Mean of Residuals 18-7
18.3.2.3 Independence of Residuals 18-7
18.3.2.4 Absence of Outliers 18-7

18.4 Fundamental Equations for Time-Series ARIMA Models 18-7

18.4.1 Identification of ARIMA (*p, d, q*) Models 18-8
18.4.1.1 Trend Components, *d:* Making the Process Stationary 18-8
18.4.1.2 Auto-Regressive Components 18-11
18.4.1.3 Moving Average Components 18-12
18.4.1.4 Mixed Models 18-13
18.4.1.5 ACFs and PACFs 18-13
18.4.2 Estimating Model Parameters 18-18
18.4.3 Diagnosing a Model 18-19
18.4.4 Computer Analysis of Small-Sample Time-Series Example 18-21

18.5 Types of Time-Series Analyses 18-29
18.5.1 Models With Seasonal Components 18-29
18.5.2 Models With Interventions 18-30
18.5.2.1 Abrupt, Permanent Effects 18-33
18.5.2.2 Abrupt, Temporary Effects 18-36
18.5.2.3 Gradual, Permanent Effects 18-36
18.5.2.4 Models With Multiple Interventions 18-40
18.5.3 Adding Continuous Variables 18-40

18.6 Some Important Issues 18-42
18.6.1 Patterns of ACFs and PACFs 18-42
18.6.2 Effect Size 18-45
18.6.3 Forecasting 18-47
18.6.4 Statistical Methods for Comparing Two Models 18-47

18.7 Complete Example of a Time-Series Analysis 18-49
18.7.1 Evaluation of Assumptions 18-49
18.7.1.1 Normality of Sampling Distributions 18-49
18.7.1.2 Homogeneity of Variance 18-49
18.7.1.3 Outliers 18-50
18.7.2 Baseline Model Identification and Estimation 18-50
18.7.3 Baseline Model Diagnosis 18-55
18.7.4 Intervention Analysis 18-55
18.7.4.1 Model Diagnosis 18-55
18.7.4.2 Model Interpretation 18-57

18.8 Comparison of Programs 18-62
18.8.1 IBM SPSS Package 18-62
18.8.2 SAS System 18-65
18.8.3 SYSTAT System 18-65

Appendix A A Skimpy Introduction to Matrix Algebra 927

A.1 The Trace of a Matrix 928

A.2 Addition or Subtraction of a Constant to a Matrix 928

A.3 Multiplication or Division of a Matrix by a Constant 928

A.4 Addition and Subtraction of Two Matrices 929

A.5 Multiplication, Transposes, and Square Roots of Matrices 929

A.6 Matrix "Division" (Inverses and Determinants) 931

A.7 Eigenvalues and Eigenvectors: Procedures for Consolidating Variance From a Matrix 933

Appendix B Research Designs for Complete Examples 937

B.1 Women's Health and Drug Study 937

B.2 Sexual Attraction Study 938

B.3 Learning Disabilities Data Bank 941

B.4 Reaction Time to Identify Figures 942

B.5 Field Studies of Noise-Induced Sleep Disturbance 942

B.6 Clinical Trial for Primary Biliary Cirrhosis 943

B.7 Impact of Seat Belt Law 943

Appendix C Statistical Tables 944

C.1 Normal Curve Areas 945

C.2 Critical Values of the t Distribution for $\alpha = .05$ and .01, Two-Tailed Test 946

C.3 Critical Values of the F Distribution 947

C.4 Critical Values of Chi Square (χ^2) 952

C.5 Critical Values for Squared Multiple Correlation (R^2) in Forward Stepwise Selection: $\alpha = .05$ 953

C.6 Critical Values for F_{MAX} (S^2_{MAX}/S^2_{MIN}) Distribution for $\alpha = .05$ and .01 955

References 956

Index 965