

CONTENTS

Preface	xxvii
---------	-------

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	4
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	20
2.1.2.3	Factorial ANOVA	20
2.1.2.4	Factorial ANCOVA	20
2.1.2.5	Hotelling's T^2	21
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	23
2.1.3	Prediction of Group Membership	23
2.1.3.1	One-Way Discriminant	23
2.1.3.2	Sequential One-Way Discriminant	24
2.1.3.3	Multiway Frequency Analysis (Logit)	24
2.1.3.4	Logistic Regression	24
2.1.3.5	Sequential Logistic Regression	25
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	26
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	27
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	53
3.3	Parameter Estimation	53
3.4	Effect Size	54
3.5	Bivariate Statistics: Correlation and Regression	56
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	96
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	118
5.2.1	Degree of Relationship	119
5.2.2	Importance of IVs	119
5.2.3	Adding IVs	119
5.2.4	Changing IVs	120
5.2.5	Contingencies among IVs	120
5.2.6	Comparing Sets of IVs	120
5.2.7	Predicting DV Scores for Members of a New Sample	120
5.2.8	Parameter Estimates	121
5.3	Limitations to Regression Analyses	121
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	124
5.3.2.4	Normality, Linearity, 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	128
5.4.1	General Linear Equations	129
5.4.2	Matrix Equations	131
5.4.3	Computer Analyses of Small-Sample Example	134

5.5	Major Types of Multiple Regression	136
5.5.1	Standard Multiple Regression	136
5.5.2	Sequential Multiple Regression	138
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	146
5.6.1.2	Sequential or Statistical Regression	146
5.6.2	Statistical Inference	146
5.6.2.1	Test for Multiple R	147
5.6.2.2	Test of Regression Components	148
5.6.2.3	Test of Added Subset of IVs	149
5.6.2.4	Confidence Limits around B and Multiple R^2	150
5.6.2.5	Comparing Two Sets of Predictors	152
5.6.3	Adjustment of R^2	153
5.6.4	Suppressor Variables	154
5.6.5	Regression Approach to ANOVA	155
5.6.6	Centering when Interactions and Powers of IVs Are Included	157
5.6.7	Mediation in Causal Sequences	159
5.7	Complete Examples of Regression Analysis	161
5.7.1	Evaluation of Assumptions	161
5.7.1.1	Ratio of Cases to IVs	161
5.7.1.2	Normality, Linearity, Homoscedasticity, and Independence of Residuals	161
5.7.1.3	Outliers	165
5.7.1.4	Multicollinearity and Singularity	167
5.7.2	Standard Multiple Regression	167
5.7.3	Sequential Regression	174
5.7.4	Example of Standard Multiple Regression with Missing Values Multiply Imputed	179
5.8	Comparison of Programs	188
5.8.1	SPSS Package	188
5.8.2	SAS System	191
5.8.3	SYSTAT System	194

6 Analysis of Covariance 195

6.1	General Purpose and Description	195
6.2	Kinds of Research Questions	198
6.2.1	Main Effects of IVs	198
6.2.2	Interactions among IVs	198
6.2.3	Specific Comparisons and Trend Analysis	199
6.2.4	Effects of Covariates	199

6.2.5	Effect Size	199
6.2.6	Parameter Estimates	199
6.3	Limitations to Analysis of Covariance	200
6.3.1	Theoretical Issues	200
6.3.2	Practical Issues	201
6.3.2.1	Unequal Sample Sizes, Missing Data, and Ratio of Cases to IVs	201
6.3.2.2	Absence of Outliers	201
6.3.2.3	Absence of Multicollinearity and Singularity	201
6.3.2.4	Normality of Sampling Distributions	202
6.3.2.5	Homogeneity of Variance	202
6.3.2.6	Linearity	202
6.3.2.7	Homogeneity of Regression	202
6.3.2.8	Reliability of Covariates	203
6.4	Fundamental Equations for Analysis of Covariance	203
6.4.1	Sums of Squares and Cross Products	204
6.4.2	Significance Test and Effect Size	208
6.4.3	Computer Analyses of Small-Sample Example	209
6.5	Some Important Issues	211
6.5.1	Choosing Covariates	211
6.5.2	Evaluation of Covariates	212
6.5.3	Test for Homogeneity of Regression	213
6.5.4	Design Complexity	213
6.5.4.1	Within-Subjects and Mixed Within-Between Designs	214
6.5.4.2	Unequal Sample Sizes	217
6.5.4.3	Specific Comparisons and Trend Analysis	218
6.5.4.4	Effect Size	221
6.5.5	Alternatives to ANCOVA	221
6.6	Complete Example of Analysis of Covariance	223
6.6.1	Evaluation of Assumptions	223
6.6.1.1	Unequal n and Missing Data	224
6.6.1.2	Normality	224
6.6.1.3	Linearity	224
6.6.1.4	Outliers	224
6.6.1.5	Multicollinearity and Singularity	227
6.6.1.6	Homogeneity of Variance	228
6.6.1.7	Homogeneity of Regression	230
6.6.1.8	Reliability of Covariates	230
6.6.2	Analysis of Covariance	230
6.6.2.1	Main Analysis	230
6.6.2.2	Evaluation of Covariates	235
6.6.2.3	Homogeneity of Regression Run	237
6.7	Comparison of Programs	240
6.7.1	SPSS Package	240

6.7.2	SAS System	240
6.7.3	SYSTAT System	240

7 Multivariate Analysis of Variance and Covariance 243

7.1	General Purpose and Description	243
7.2	Kinds of Research Questions	247
7.2.1	Main Effects of IVs	247
7.2.2	Interactions among IVs	247
7.2.3	Importance of DVs	247
7.2.4	Parameter Estimates	248
7.2.5	Specific Comparisons and Trend Analysis	248
7.2.6	Effect Size	248
7.2.7	Effects of Covariates	248
7.2.8	Repeated-Measures Analysis of Variance	249
7.3	Limitations to Multivariate Analysis of Variance and Covariance	249
7.3.1	Theoretical Issues	249
7.3.2	Practical Issues	250
7.3.2.1	Unequal Sample Sizes, Missing Data, and Power	250
7.3.2.2	Multivariate Normality	251
7.3.2.3	Absence of Outliers	251
7.3.2.4	Homogeneity of Variance-Covariance Matrices	251
7.3.2.5	Linearity	252
7.3.2.6	Homogeneity of Regression	252
7.3.2.7	Reliability of Covariates	253
7.3.2.8	Absence of Multicollinearity and Singularity	253
7.4	Fundamental Equations for Multivariate Analysis of Variance and Covariance	253
7.4.1	Multivariate Analysis of Variance	253
7.4.2	Computer Analyses of Small-Sample Example	261
7.4.3	Multivariate Analysis of Covariance	264
7.5	Some Important Issues	268
7.5.1	MANOVA vs. ANOVAs	268
7.5.2	Criteria for Statistical Inference	269
7.5.3	Assessing DVs	270
7.5.3.1	Univariate F	270
7.5.3.2	Roy-Bargmann Stepdown Analysis	271
7.5.3.3	Using Discriminant Analysis	272
7.5.3.4	Choosing among Strategies for Assessing DVs	273
7.5.4	Specific Comparisons and Trend Analysis	273
7.5.5	Design Complexity	274
7.5.5.1	Within-Subjects and Between-Within Designs	274
7.5.5.2	Unequal Sample Sizes	276

7.6 Complete Examples of Multivariate Analysis of Variance and Covariance 277

- 7.6.1 Evaluation of Assumptions 277
 - 7.6.1.1 Unequal Sample Sizes and Missing Data 277
 - 7.6.1.2 Multivariate Normality 279
 - 7.6.1.3 Linearity 279
 - 7.6.1.4 Outliers 279
 - 7.6.1.5 Homogeneity of Variance-Covariance Matrices 280
 - 7.6.1.6 Homogeneity of Regression 281
 - 7.6.1.7 Reliability of Covariates 284
 - 7.6.1.8 Multicollinearity and Singularity 285
- 7.6.2 Multivariate Analysis of Variance 285
- 7.6.3 Multivariate Analysis of Covariance 296
 - 7.6.3.1 Assessing Covariates 296
 - 7.6.3.2 Assessing DVs 296

7.7 Comparison of Programs 307

- 7.7.1 SPSS Package 307
- 7.7.2 SAS System 310
- 7.7.3 SYSTAT System 310

8 Profile Analysis: The Multivariate Approach to Repeated Measures 311

8.1 General Purpose and Description 311

8.2 Kinds of Research Questions 312

- 8.2.1 Parallelism of Profiles 312
- 8.2.2 Overall Difference among Groups 313
- 8.2.3 Flatness of Profiles 313
- 8.2.4 Contrasts Following Profile Analysis 313
- 8.2.5 Parameter Estimates 313
- 8.2.6 Effect Size 314

8.3 Limitations to Profile Analysis 314

- 8.3.1 Theoretical Issues 314
- 8.3.2 Practical Issues 315
 - 8.3.2.1 Sample Size, Missing Data, and Power 315
 - 8.3.2.2 Multivariate Normality 315
 - 8.3.2.3 Absence of Outliers 315
 - 8.3.2.4 Homogeneity of Variance-Covariance Matrices 315
 - 8.3.2.5 Linearity 316
 - 8.3.2.6 Absence of Multicollinearity and Singularity 316

8.4 Fundamental Equations for Profile Analysis 316

- 8.4.1 Differences in Levels 316
- 8.4.2 Parallelism 318

8.4.3	Flatness	321
8.4.4	Computer Analyses of Small-Sample Example	323
8.5	Some Important Issues	329
8.5.1	Univariate vs. Multivariate Approach to Repeated Measures	329
8.5.2	Contrasts in Profile Analysis	331
8.5.2.1	Parallelism and Flatness Significant, Levels Not Significant (Simple-effects Analysis)	333
8.5.2.2	Parallelism and Levels Significant, Flatness Not Significant (Simple-effects Analysis)	336
8.5.2.3	Parallelism, Levels, and Flatness Significant (Interaction Contrasts)	339
8.5.2.4	Only Parallelism Significant	339
8.5.3	Doubly-Multivariate Designs	339
8.5.4	Classifying Profiles	345
8.5.5	Imputation of Missing Values	345
8.6	Complete Examples of Profile Analysis	346
8.6.1	Profile Analysis of Subscales of the WISC	346
8.6.1.1	Evaluation of Assumptions	346
8.6.1.2	Profile Analysis	351
8.6.2	Doubly-Multivariate Analysis of Reaction Time	360
8.6.2.1	Evaluation of Assumptions	360
8.6.2.2	Doubly-Multivariate Analysis of Slope and Intercept	363
8.7	Comparison of Programs	371
8.7.1	SPSS Package	373
8.7.2	SAS System	373
8.7.3	SYSTAT System	374

9 Discriminant Analysis 375

9.1	General Purpose and Description	375
9.2	Kinds of Research Questions	378
9.2.1	Significance of Prediction	378
9.2.2	Number of Significant Discriminant Functions	378
9.2.3	Dimensions of Discrimination	379
9.2.4	Classification Functions	379
9.2.5	Adequacy of Classification	379
9.2.6	Effect Size	379
9.2.7	Importance of Predictor Variables	380
9.2.8	Significance of Prediction with Covariates	380
9.2.9	Estimation of Group Means	380
9.3	Limitations to Discriminant Analysis	381
9.3.1	Theoretical Issues	381

9.3.2	Practical Issues	381
9.3.2.1	Unequal Sample Sizes, Missing Data, and Power	381
9.3.2.2	Multivariate Normality	382
9.3.2.3	Absence of Outliers	382
9.3.2.4	Homogeneity of Variance-Covariance Matrices	382
9.3.2.5	Linearity	383
9.3.2.6	Absence of Multicollinearity and Singularity	383
9.4	Fundamental Equations for Discriminant Analysis	384
9.4.1	Derivation and Test of Discriminant Functions	384
9.4.2	Classification	387
9.4.3	Computer Analyses of Small-Sample Example	389
9.5	Types of Discriminant Function Analyses	395
9.5.1	Direct Discriminant Analysis	395
9.5.2	Sequential Discriminant Analysis	396
9.5.3	Stepwise (Statistical) Discriminant Analysis	396
9.6	Some Important Issues	397
9.6.1	Statistical Inference	397
9.6.1.1	Criteria for Overall Statistical Significance	397
9.6.1.2	Stepping Methods	397
9.6.2	Number of Discriminant Functions	398
9.6.3	Interpreting Discriminant Functions	398
9.6.3.1	Discriminant Function Plots	398
9.6.3.2	Structure Matrix of Loadings	400
9.6.4	Evaluating Predictor Variables	401
9.6.5	Effect Size	402
9.6.6	Design Complexity: Factorial Designs	403
9.6.7	Use of Classification Procedures	404
9.6.7.1	Cross-Validation and New Cases	405
9.6.7.2	Jackknifed Classification	405
9.6.7.3	Evaluating Improvement in Classification	405
9.7	Complete Example of Discriminant Analysis	407
9.7.1	Evaluation of Assumptions	407
9.7.1.1	Unequal Sample Sizes and Missing Data	407
9.7.1.2	Multivariate Normality	408
9.7.1.3	Linearity	408
9.7.1.4	Outliers	408
9.7.1.5	Homogeneity of Variance-Covariance Matrices	411
9.7.1.6	Multicollinearity and Singularity	411
9.7.2	Direct Discriminant Analysis	412
9.8	Comparison of Programs	430
9.8.1	SPSS Package	430
9.8.2	SAS System	430
9.8.3	SYSTAT System	436

10	Logistic Regression	437
10.1	General Purpose and Description	437
10.2	Kinds of Research Questions	439
10.2.1	Prediction of Group Membership or Outcome	439
10.2.2	Importance of Predictors	439
10.2.3	Interactions among Predictors	440
10.2.4	Parameter Estimates	440
10.2.5	Classification of Cases	440
10.2.6	Significance of Prediction with Covariates	440
10.2.7	Effect Size	441
10.3	Limitations to Logistic Regression Analysis	441
10.3.1	Theoretical Issues	441
10.3.2	Practical Issues	442
10.3.2.1	Ratio of Cases to Variables	442
10.3.2.2	Adequacy of Expected Frequencies and Power	442
10.3.2.3	Linearity in the Logit	443
10.3.2.4	Absence of Multicollinearity	443
10.3.2.5	Absence of Outliers in the Solution	443
10.3.2.6	Independence of Errors	443
10.4	Fundamental Equations for Logistic Regression	444
10.4.1	Testing and Interpreting Coefficients	445
10.4.2	Goodness-of-Fit	446
10.4.3	Comparing Models	448
10.4.4	Interpretation and Analysis of Residuals	448
10.4.5	Computer Analyses of Small-Sample Example	449
10.5	Types of Logistic Regression	453
10.5.1	Direct Logistic Regression	454
10.5.2	Sequential Logistic Regression	454
10.5.3	Statistical (Stepwise) Logistic Regression	454
10.5.4	Probit and Other Analyses	456
10.6	Some Important Issues	457
10.6.1	Statistical Inference	457
10.6.1.1	Assessing Goodness-of-Fit of Models	457
10.6.1.2	Tests of Individual Variables	459
10.6.2	Effect Size for a Model	460
10.6.3	Interpretation of Coefficients Using Odds	461
10.6.4	Coding Outcome and Predictor Categories	464
10.6.5	Number and Type of Outcome Categories	464
10.6.6	Classification of Cases	468
10.6.7	Hierarchical and Nonhierarchical Analysis	468

10.6.8	Importance of Predictors	469
10.6.9	Logistic Regression for Matched Groups	469
10.7	Complete Examples of Logistic Regression	469
10.7.1	Evaluation of Limitations	470
10.7.1.1	Ratio of Cases to Variables and Missing Data	470
10.7.1.2	Multicollinearity	473
10.7.1.3	Outliers in the Solution	474
10.7.2	Direct Logistic Regression with Two-Category Outcome and Continuous Predictors	474
10.7.2.1	Limitation: Linearity in the Logit	474
10.7.2.2	Direct Logistic Regression with Two-Category Outcome	474
10.7.3	Sequential Logistic Regression with Three Categories of Outcome	481
10.7.3.1	Limitations of Multinomial Logistic Regression	481
10.7.3.2	Sequential Multinomial Logistic Regression	481
10.8	Comparisons of Programs	499
10.8.1	SPSS Package	499
10.8.2	SAS System	504
10.8.3	SYSTAT System	504

11 Survival/Failure Analysis 506

11.1	General Purpose and Description	506
11.2	Kinds of Research Questions	507
11.2.1	Proportions Surviving at Various Times	507
11.2.2	Group Differences in Survival	508
11.2.3	Survival Time with Covariates	508
11.2.3.1	Treatment Effects	508
11.2.3.2	Importance of Covariates	508
11.2.3.3	Parameter Estimates	508
11.2.3.4	Contingencies among Covariates	508
11.2.3.5	Effect Size and Power	509
11.3	Limitations to Survival Analysis	509
11.3.1	Theoretical Issues	509
11.3.2	Practical Issues	509
11.3.2.1	Sample Size and Missing Data	509
11.3.2.2	Normality of Sampling Distributions, Linearity, and Homoscedasticity	510
11.3.2.3	Absence of Outliers	510
11.3.2.4	Differences between Withdrawn and Remaining Cases	510
11.3.2.5	Change in Survival Conditions over Time	510
11.3.2.6	Proportionality of Hazards	510
11.3.2.7	Absence of Multicollinearity	510

11.4	Fundamental Equations for Survival Analysis	511
11.4.1	Life Tables	511
11.4.2	Standard Error of Cumulative Proportion Surviving	513
11.4.3	Hazard and Density Functions	514
11.4.4	Plot of Life Tables	515
11.4.5	Test for Group Differences	515
11.4.6	Computer Analyses of Small-Sample Example	517
11.5	Types of Survival Analyses	524
11.5.1	Actuarial and Product-Limit Life Tables and Survivor Functions	524
11.5.2	Prediction of Group Survival Times from Covariates	524
11.5.2.1	Direct, Sequential, and Statistical Analysis	527
11.5.2.2	Cox Proportional-Hazards Model	527
11.5.2.3	Accelerated Failure-Time Models	529
11.5.2.4	Choosing a Method	535
11.6	Some Important Issues	535
11.6.1	Proportionality of Hazards	535
11.6.2	Censored Data	537
11.6.2.1	Right-Censored Data	537
11.6.2.2	Other Forms of Censoring	537
11.6.3	Effect Size and Power	538
11.6.4	Statistical Criteria	539
11.6.4.1	Test Statistics for Group Differences in Survival Functions	539
11.6.4.2	Test Statistics for Prediction from Covariates	540
11.6.5	Predicting Survival Rate	540
11.6.5.1	Regression Coefficients (Parameter Estimates)	540
11.6.5.2	Odds Ratios	540
11.6.5.3	Expected Survival Rates	541
11.7	Complete Example of Survival Analysis	541
11.7.1	Evaluation of Assumptions	543
11.7.1.1	Accuracy of Input, Adequacy of Sample Size, Missing Data, and Distributions	543
11.7.1.2	Outliers	545
11.7.1.3	Differences between Withdrawn and Remaining Cases	549
11.7.1.4	Change in Survival Experience over Time	549
11.7.1.5	Proportionality of Hazards	549
11.7.1.6	Multicollinearity	551
11.7.2	Cox Regression Survival Analysis	551
11.7.2.1	Effect of Drug Treatment	552
11.7.2.2	Evaluation of Other Covariates	552
11.8	Comparison of Programs	559
11.8.1	SAS System	559
11.8.2	SPSS Package	559
11.8.3	SYSTAT System	566

12 Canonical Correlation 567

- 12.1 General Purpose and Description 567
- 12.2 Kinds of Research Questions 568
 - 12.2.1 Number of Canonical Variate Pairs 568
 - 12.2.2 Interpretation of Canonical Variates 569
 - 12.2.3 Importance of Canonical Variates 569
 - 12.2.4 Canonical Variate Scores 569
- 12.3 Limitations 569
 - 12.3.1 Theoretical Limitations 569
 - 12.3.2 Practical Issues 570
 - 12.3.2.1 Ratio of Cases to IVs 570
 - 12.3.2.2 Normality, Linearity, and Homoscedasticity 570
 - 12.3.2.3 Missing Data 571
 - 12.3.2.4 Absence of Outliers 571
 - 12.3.2.5 Absence of Multicollinearity and Singularity 571
- 12.4 Fundamental Equations for Canonical Correlation 572
 - 12.4.1 Eigenvalues and Eigenvectors 573
 - 12.4.2 Matrix Equations 575
 - 12.4.3 Proportions of Variance Extracted 579
 - 12.4.4 Computer Analyses of Small-Sample Example 580
- 12.5 Some Important Issues 586
 - 12.5.1 Importance of Canonical Variates 586
 - 12.5.2 Interpretation of Canonical Variates 587
- 12.6 Complete Example of Canonical Correlation 587
 - 12.6.1 Evaluation of Assumptions 588
 - 12.6.1.1 Missing Data 588
 - 12.6.1.2 Normality, Linearity, and Homoscedasticity 588
 - 12.6.1.3 Outliers 591
 - 12.6.1.4 Multicollinearity and Singularity 595
 - 12.6.2 Canonical Correlation 595
- 12.7 Comparison of Programs 604
 - 12.7.1 SAS System 604
 - 12.7.2 SPSS Package 604
 - 12.7.3 SYSTAT System 606

13 Principal Components and Factor Analysis 607

- 13.1 General Purpose and Description 607
- 13.2 Kinds of Research Questions 610
 - 13.2.1 Number of Factors 610

13.2.2	Nature of Factors	611
13.2.3	Importance of Solutions and Factors	611
13.2.4	Testing Theory in FA	611
13.2.5	Estimating Scores on Factors	611
13.3	Limitations	611
13.3.1	Theoretical Issues	611
13.3.2	Practical Issues	612
13.3.2.1	Sample Size and Missing Data	613
13.3.2.2	Normality	613
13.3.2.3	Linearity	613
13.3.2.4	Absence of Outliers among Cases	613
13.3.2.5	Absence of Multicollinearity and Singularity	614
13.3.2.6	Factorability of $\mathbf{R}$	614
13.3.2.7	Absence of Outliers among Variables	614
13.4	Fundamental Equations for Factor Analysis	615
13.4.1	Extraction	616
13.4.2	Orthogonal Rotation	620
13.4.3	Communalities, Variance, and Covariance	621
13.4.4	Factor Scores	622
13.4.5	Oblique Rotation	625
13.4.6	Computer Analyses of Small-Sample Example	628
13.5	Major Types of Factor Analyses	633
13.5.1	Factor Extraction Techniques	633
13.5.1.1	PCA vs. FA	634
13.5.1.2	Principal Components	635
13.5.1.3	Principal Factors	636
13.5.1.4	Image Factor Extraction	636
13.5.1.5	Maximum Likelihood Factor Extraction	636
13.5.1.6	Unweighted Least Squares Factoring	636
13.5.1.7	Generalized (Weighted) Least Squares Factoring	637
13.5.1.8	Alpha Factoring	637
13.5.2	Rotation	637
13.5.2.1	Orthogonal Rotation	638
13.5.2.2	Oblique Rotation	638
13.5.2.3	Geometric Interpretation	640
13.5.3	Some Practical Recommendations	642
13.6	Some Important Issues	643
13.6.1	Estimates of Communalities	643
13.6.2	Adequacy of Extraction and Number of Factors	644
13.6.3	Adequacy of Rotation and Simple Structure	646
13.6.4	Importance and Internal Consistency of Factors	647
13.6.5	Interpretation of Factors	649
13.6.6	Factor Scores	650
13.6.7	Comparisons among Solutions and Groups	651

13.7	Complete Example of FA	651
13.7.1	Evaluation of Limitations	652
13.7.1.1	Sample Size and Missing Data	652
13.7.1.2	Normality	652
13.7.1.3	Linearity	652
13.7.1.4	Outliers	652
13.7.1.5	Multicollinearity and Singularity	657
13.7.1.6	Outliers among Variables	657
13.7.2	Principal Factors Extraction with Varimax Rotation	657
13.8	Comparison of Programs	671
13.8.1	SPSS Package	674
13.8.2	SAS System	675
13.8.3	SYSTAT System	675
14	Structural Equation Modeling	676
	by Jodie B. Ullman	
14.1	General Purpose and Description	676
14.2	Kinds of Research Questions	680
14.2.1	Adequacy of the Model	680
14.2.2	Testing Theory	680
14.2.3	Amount of Variance in the Variables Accounted for by the Factors	680
14.2.4	Reliability of the Indicators	680
14.2.5	Parameter Estimates	680
14.2.6	Intervening Variables	681
14.2.7	Group Differences	681
14.2.8	Longitudinal Differences	681
14.2.9	Multilevel Modeling	681
14.3	Limitations to Structural Equation Modeling	682
14.3.1	Theoretical Issues	682
14.3.2	Practical Issues	682
14.3.2.1	Sample Size and Missing Data	682
14.3.2.2	Multivariate Normality and Absence of Outliers	683
14.3.2.3	Linearity	683
14.3.2.4	Absence of Multicollinearity and Singularity	683
14.3.2.5	Residuals	684
14.4	Fundamental Equations for Structural Equations Modeling	684
14.4.1	Covariance Algebra	684
14.4.2	Model Hypotheses	686
14.4.3	Model Specification	688
14.4.4	Model Estimation	690
14.4.5	Model Evaluation	694
14.4.6	Computer Analysis of Small-Sample Example	696

14.5	Some Important Issues	709
14.5.1	Model Identification	709
14.5.2	Estimation Techniques	713
14.5.2.1	Estimation Methods and Sample Size	714
14.5.2.2	Estimation Methods and Nonnormality	714
14.5.2.3	Estimation Methods and Dependence	715
14.5.2.4	Some Recommendations for Choice of Estimation Method	715
14.5.3	Assessing the Fit of the Model	715
14.5.3.1	Comparative Fit Indices	716
14.5.3.2	Absolute Fit Index	718
14.5.3.3	Indices of Proportion of Variance Accounted	718
14.5.3.4	Degree of Parsimony Fit Indices	719
14.5.3.5	Residual-Based Fit Indices	720
14.5.3.6	Choosing among Fit Indices	720
14.5.4	Model Modification	721
14.5.4.1	Chi-Square Difference Test	721
14.5.4.2	Lagrange Multiplier (LM) Test	721
14.5.4.3	Wald Test	723
14.5.4.4	Some Caveats and Hints on Model Modification	728
14.5.5	Reliability and Proportion of Variance	728
14.5.6	Discrete and Ordinal Data	729
14.5.7	Multiple Group Models	730
14.5.8	Mean and Covariance Structure Models	731
14.6	Complete Examples of Structural Equation Modeling Analysis	732
14.6.1	Confirmatory Factor Analysis of the WISC	732
14.6.1.1	Model Specification for CFA	732
14.6.1.2	Evaluation of Assumptions for CFA	733
14.6.1.3	CFA Model Estimation and Preliminary Evaluation	734
14.6.1.4	Model Modification	743
14.6.2	SEM of Health Data	750
14.6.2.1	SEM Model Specification	750
14.6.2.2	Evaluation of Assumptions for SEM	751
14.6.2.3	SEM Model Estimation and Preliminary Evaluation	755
14.6.2.4	Model Modification	759
14.7	Comparison of Programs	773
14.7.1	EQS	773
14.7.2	LISREL	773
14.7.3	AMOS	780
14.7.4	SAS System	780

15 Multilevel Linear Modeling 781

15.1	General Purpose and Description	781
-------------	--	------------

15.2	Kinds of Research Questions	784
15.2.1	Group Differences in Means	784
15.2.2	Group Differences in Slopes	784
15.2.3	Cross-Level Interactions	785
15.2.4	Meta-Analysis	785
15.2.5	Relative Strength of Predictors at Various Levels	785
15.2.6	Individual and Group Structure	785
15.2.7	Path Analysis at Individual and Group Levels	786
15.2.8	Analysis of Longitudinal Data	786
15.2.9	Multilevel Logistic Regression	786
15.2.10	Multiple Response Analysis	786
15.3	Limitations to Multilevel Linear Modeling	786
15.3.1	Theoretical Issues	786
15.3.2	Practical Issues	787
15.3.2.1	Sample Size, Unequal- n , and Missing Data	787
15.3.2.2	Independence of Errors	788
15.3.2.3	Absence of Multicollinearity and Singularity	789
15.4	Fundamental Equations	789
15.4.1	Intercepts-Only Model	792
15.4.1.1	The Intercepts-Only Model: Level-1 Equation	793
15.4.1.2	The Intercepts-Only Model: Level-2 Equation	793
15.4.1.3	Computer Analysis of Intercepts-only Model	794
15.4.2	Model with a First-Level Predictor	799
15.4.2.1	Level-1 Equation for a Model with a Level-1 Predictor	799
15.4.2.2	Level-2 Equations for a Model with a Level-1 Predictor	801
15.4.2.3	Computer Analysis of a Model with a Level-1 Predictor	802
15.4.3	Model with Predictors at First and Second Levels	807
15.4.3.1	Level-1 Equation for Model with Predictors at Both Levels	807
15.4.3.2	Level-2 Equations for Model with Predictors at Both Levels	807
15.4.3.3	Computer Analyses of Model with Predictors at First and Second Levels	808
15.5	Types of MLM	814
15.5.1	Repeated Measures	814
15.5.2	Higher-Order MLM	819
15.5.3	Latent Variables	819
15.5.4	Nonnormal Outcome Variables	820
15.5.5	Multiple Response Models	821
15.6	Some Important Issues	822
15.6.1	Intraclass Correlation	822
15.6.2	Centering Predictors and Changes in Their Interpretations	823
15.6.3	Interactions	826
15.6.4	Random and Fixed Intercepts and Slopes	826

15.6.5	Statistical Inference	830
15.6.5.1	Assessing Models	830
15.6.5.2	Tests of Individual Effects	831
15.6.6	Effect Size	832
15.6.7	Estimation Techniques and Convergence Problems	833
15.6.8	Exploratory Model Building	834
15.7	Complete Example of MLM	835
15.7.1	Evaluation of Assumptions	835
15.7.1.1	Sample Sizes, Missing Data, and Distributions	835
15.7.1.2	Outliers	838
15.7.1.3	Multicollinearity and Singularity	839
15.7.1.4	Independence of Errors: Intraclass Correlations	839
15.7.2	Multilevel Modeling	840
15.8	Comparison of Programs	852
15.8.1	SAS System	852
15.8.2	SPSS Package	856
15.8.3	HLM Program	856
15.8.4	MLwiN Program	857
15.8.5	SYSTAT System	857
16	Multiway Frequency Analysis	858
16.1	General Purpose and Description	858
16.2	Kinds of Research Questions	859
16.2.1	Associations among Variables	859
16.2.2	Effect on a Dependent Variable	860
16.2.3	Parameter Estimates	860
16.2.4	Importance of Effects	860
16.2.5	Effect Size	860
16.2.6	Specific Comparisons and Trend Analysis	860
16.3	Limitations to Multiway Frequency Analysis	861
16.3.1	Theoretical Issues	861
16.3.2	Practical Issues	861
16.3.2.1	Independence	861
16.3.2.2	Ratio of Cases to Variables	861
16.3.2.3	Adequacy of Expected Frequencies	862
16.3.2.4	Absence of Outliers in the Solution	863
16.4	Fundamental Equations for Multiway Frequency Analysis	863
16.4.1	Screening for Effects	864
16.4.1.1	Total Effect	865
16.4.1.2	First-Order Effects	866
16.4.1.3	Second-Order Effects	867
16.4.1.4	Third-Order Effect	871

16.4.2	Modeling	871
16.4.3	Evaluation and Interpretation	874
16.4.3.1	Residuals	874
16.4.3.2	Parameter Estimates	874
16.4.4	Computer Analyses of Small-Sample Example	880
16.5	Some Important Issues	887
16.5.1	Hierarchical and Nonhierarchical Models	887
16.5.2	Statistical Criteria	888
16.5.2.1	Tests of Models	888
16.5.2.2	Tests of Individual Effects	888
16.5.3	Strategies for Choosing a Model	889
16.5.3.1	SPSS HILOGLINEAR (Hierarchical)	889
16.5.3.2	SPSS GENLOG (General Log-Linear)	889
16.5.3.3	SAS CATMOD and SPSS LOGLINEAR (General Log-Linear)	890
16.6	Complete Example of Multiway Frequency Analysis	890
16.6.1	Evaluation of Assumptions: Adequacy of Expected Frequencies	890
16.6.2	Hierarchical Log-Linear Analysis	891
16.6.2.1	Preliminary Model Screening	891
16.6.2.2	Stepwise Model Selection	893
16.6.2.3	Adequacy of Fit	895
16.6.2.4	Interpretation of the Selected Model	901
16.7	Comparison of Programs	908
16.7.1	SPSS Package	911
16.7.2	SAS System	912
16.7.3	SYSTAT System	912
17	An Overview of the General Linear Model	913
17.1	Linearity and the General Linear Model	913
17.2	Bivariate to Multivariate Statistics and Overview of Techniques	913
17.2.1	Bivariate Form	913
17.2.2	Simple Multivariate Form	914
17.2.3	Full Multivariate Form	917
17.3	Alternative Research Strategies	918
18	Time-Series Analysis (available online at www.ablongman.com/tabachnick5e)	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 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	18-16
18.4.3	Diagnosing a Model	18-19
18.4.4	Computer Analysis of Small-Sample Time-Series Example	18-19
18.5	Types of Time-Series Analyses	18-27
18.5.1	Models with Seasonal Components	18-27
18.5.2	Models with Interventions	18-30
18.5.2.1	Abrupt, Permanent Effects	18-32
18.5.2.2	Abrupt, Temporary Effects	18-32
18.5.2.3	Gradual, Permanent Effects	18-38
18.5.2.4	Models with Multiple Interventions	18-38
18.5.3	Adding Continuous Variables	18-38
18.6	Some Important Issues	18-41
18.6.1	Patterns of ACFs and PACFs	18-41
18.6.2	Effect Size	18-44
18.6.3	Forecasting	18-45
18.6.4	Statistical Methods for Comparing Two Models	18-45
18.7	Complete Example of a Time-Series Analysis	18-47
18.7.1	Evaluation of Assumptions	18-48
18.7.1.1	Normality of Sampling Distributions	18-48
18.7.1.2	Homogeneity of Variance	18-48
18.7.1.3	Outliers	18-48

18.7.2	Baseline Model Identification and Estimation	18-48
18.7.3	Baseline Model Diagnosis	18-49
18.7.4	Intervention Analysis	18-55
18.7.4.1	Model Diagnosis	18-55
18.7.4.2	Model Interpretation	18-56
18.8	Comparison of Programs	18-60
18.8.1	SPSS Package	18-61
18.8.2	SAS System	18-61
18.8.3	SYSTAT System	18-61

Appendix A A Skimpy Introduction to Matrix Algebra 924

A.1	The Trace of a Matrix	925
A.2	Addition or Subtraction of a Constant to a Matrix	925
A.3	Multiplication or Division of a Matrix by a Constant	925
A.4	Addition and Subtraction of Two Matrices	926
A.5	Multiplication, Transposes, and Square Roots of Matrices	927
A.6	Matrix "Division" (Inverses and Determinants)	929
A.7	Eigenvalues and Eigenvectors: Procedures for Consolidating Variance from a Matrix	930

Appendix B Research Designs for Complete Examples 934

B.1	Women's Health and Drug Study	934
B.2	Sexual Attraction Study	935
B.3	Learning Disabilities Data Bank	938
B.4	Reaction Time to Identify Figures	939
B.5	Field Studies of Noise-Induced Sleep Disturbance	939
B.6	Clinical Trial for Primary Biliary Cirrhosis	940
B.7	Impact of Seat Belt Law	940

Appendix C Statistical Tables 941

C.1	Normal Curve Areas	942
C.2	Critical Values of the t Distribution for $\alpha = .05$ and $.01$, Two-Tailed Test	943

PREFACE

C.3 Critical Values of the F Distribution 944

C.4 Critical Values of Chi Square (χ^2) 949

C.5 Critical Values for Squared Multiple Correlation (R^2) in Forward Stepwise Selection 950

C.6 Critical Values for $F_{\text{MAX}} (S^2_{\text{MAX}}/S^2_{\text{MIN}})$ Distribution for $\alpha = .05$ and $.01$ 952

References 953

Index 963

Over my threatened, and I'm sure you're considering putting the book on a diet. We've asked only one chapter this time for dealing with missing data (Chapter 4). Otherwise, we've mostly streamlined and said good-bye to some old friends. We've forsaken the Time-Series Analysis chapter in the text, but you'll be able to download it from the publisher's web site at www.ablongman.com/tabachnick5e. Another old friend is SYSTAT. We still love the program, however, for its right-to-the-point analyses and terrific graphics, and are pleased that most of the graphics have been incorporated into SPSS. Although absent from demonstrations, features of SYSTAT, and any other programs we've used are compared. We've changed the order of some chapters: canonical correlation seemed rather difficult to appear as early as it did, and survival analysis seemed to want to smuggle up to logistic regression. Actually, the order doesn't seem to matter much; perusal of syllabi on the Web convinces us that professors feel free to present chapters in any order they choose—and that's fine with us.

Multilevel linear modeling (MLM) seems to have taken the world by storm; how did we ever live without it? Real life is hierarchical—students come to us within classrooms, teachers work within different schools, patients share wards and nursing staff, and audiences attend different performances. We hardly ever go on break these groups apart for research purposes, so we have to deal with intact groups and all their shared experiences. MLM lets us do this without violating all of the statistical assumptions we learned to know and hate. Now that SAS and SPSS can deal with these models, we're ready to tackle the real world. Hence, a new chapter.

SAS and SPSS also now offer reasonable ways to impute missing data through multiple-imputation techniques and fully assess missing data patterns, respectively. We expanded Chapter 4 to demonstrate these enhancements. SPSS and SAS keep adding goodies, which we'll try to show off. As before, we adopt our syntax from Windows menus whenever possible, and all of our data sets are available on the book's web page (www.ablongman.com/tabachnick5e). We've also paid more attention to effect sizes and, especially, confidence intervals around effect sizes. Michael Smithson of the Australian National University has kindly given us permission to include some nifty SPSS and SAS syntax and data files in our web page downloads. Jim Steiger and Rachel Fouladi have graciously given us permission to include their DOS program that finds confidence intervals around η^2 .

One thing we'll never change is our practical bent, focusing on the benefits and drawbacks of applications of a technique to a data set—when, why, and how to do it. The math is wonderful, but we suggest (but don't insist) that students follow along through section four of each chapter using readily available software for matrix manipulations or spreadsheets. Understanding the math is not enough to insure appropriate analysis of data. And, as we've said before, they really are able to apply the techniques without a great deal of fuss. The math of section four. Our small-sample examples remain silly; alas, our hefty data sets are not. As for our most recent reviewers, kindly provided by our publisher, we had one who was picking out beds too hard, too soft, and just right. So we've not changed the book's format at all.

Some extremely helpful advice was offered by John Kline of the University of Missouri-Columbia and Jeremy Jewell of Southern Illinois University-Edwardsville. We also