Contents

Preface xiii

Preface to the second edition xvii

Preface to the first edition xix

Part One Preparation for Analysis 1

1 What is multivariate analysis? 3
1.1 How is multivariate analysis defined? 3
1.2 Examples of studies in which multivariate analysis is useful 3
1.3 Multivariate analyses discussed in this book 6
1.4 Organization and content of the book 9
References 11

2 Characterizing data for future analyses 12
2.1 Variables: their definition, classification and use 12
2.2 Defining statistical variables 12
2.3 How variables are classified: Stevens's classification system 13
2.4 How variables are used in data analysis 16
2.5 Examples of classifying variables 17
2.6 Other characteristics of data 18
Summary 18
References 19
Further reading 19
Problems 19

3 Preparing for data analysis 21
3.1 Processing the data so they can be analyzed 21
3.2 Choice of computer for statistical analysis 22
3.3 Choice of a statistical package 23
3.4 Techniques for data entry 28

3.5 Data management for statistics 34
3.6 Data example: Los Angeles depression study 40
Summary 43
References 45
Further reading 46
Problems 46

4 Data screening and data transformation 48
4.1 Making transformations and assessing normality and independence 48
4.2 Common transformations 48
4.3 Assessing the need for and selecting a transformation 54
4.4 Assessing independence 64
Summary 67
References 67
Further reading 68
Problems 68

5 Selecting appropriate analyses 71
5.1 Which analyses? 71
5.2 Why selection of analyses is often difficult 71
5.3 Appropriate statistical measures under Stevens's classification 72
5.4 Appropriate multivariate analyses under Stevens's classification 76
Summary 79
References 79
Further reading 80
Problems 80

Part Two Applied Regression Analysis 83

6 Simple linear regression and correlation 85
6.1 Using linear regression and correlation to examine the relationship between two variables 85
6.2 When are regression and correlation used? 85
6.3 Data example 86
6.4 Description of methods of regression: fixed-X case 88
6.5 Description of methods of regression and correlation: variable-X case 93
6.6 Interpretation of results: fixed-X case 94
6.7 Interpretation of results: variable-X case 96
6.8 Further examination of computer output 100

6.9 Robustness and transformations for regression analysis 108
6.10 Other options in computer programs 111
6.11 Special applications of regression 112
6.12 Discussion of computer programs 115
6.13 What to watch out for 117
Summary 118
References 118
Further reading 120
Problems 121

7 Multiple regression and correlation 124
7.1 Using multiple linear regression to examine the relationship between one dependent variable and multiple independent variables 124
7.2 When are multiple regression and correlation used? 125
7.3 Data example 125
7.4 Description of techniques: fixed-X case 128
7.5 Description of techniques: variable-X case 130
7.6 How to interpret the results: fixed-X case 137
7.7 How to interpret the results: variable-X case 140
7.8 Residual analysis and transformations 143
7.9 Other options in computer programs 148
7.10 Discussion of computer programs 154
7.11 What to watch out for 157
Summary 160
References 160
Further reading 161
Problems 162

8 Variable selection in regression analysis 166
8.1 Using variable selection techniques in multiple regression analysis 166
8.2 When are variable selection methods used? 166
8.3 Data example 167
8.4 Criteria for variable selection 170
8.5 A general F test 173
8.6 Stepwise regression 175
8.7 Subset regression 181
8.8 Discussion of computer programs 185
8.9 Discussion and extensions 187
8.10 What to watch out for 191
Summary 193
References 193

Further reading 194
Problems 194

9 Special regression topics 197
9.1 Special topics in regression analysis 197
9.2 Missing values in regression analysis 197
9.3 Dummy variables 202
9.4 Constraints on parameters 209
9.5 Methods for obtaining a regression equation when multicollinearity is present 212
9.6 Ridge regression 214
Summary 219
References 220
Further reading 221
Problems 221

Part Three Multivariate Analysis 225

10 Canonical correlation analysis 227
10.1 Using canonical correlation analysis to analyze two sets of variables 227
10.2 When is canonical correlation analysis used? 227
10.3 Data example 228
10.4 Basic concepts of canonical correlation 229
10.5 Other topics related to canonical correlation 234
10.6 Discussion of computer programs 237
10.7 What to watch out for 239
Summary 240
References 241
Further reading 241
Problems 241

11 Discriminant analysis 243
11.1 Using discriminant analysis to classify cases 243
11.2 When is discriminant analysis used? 244
11.3 Data example 245
11.4 Basic concepts of classification 246
11.5 Theoretical background 253
11.6 Interpretation 255
11.7 Adjusting the value of the dividing point 259
11.8 How good is the discriminant function? 262
11.9 Testing for the contributions of classification variables 265
11.10 Variable selection 266

11.11 Classification into more than two groups 267
11.12 Use of canonical correlation in discriminant function analysis 269
11.13 Discussion of computer programs 272
11.14 What to watch out for 275
Summary 276
References 277
Further reading 277
Problems 278

12 Logistic regression 281
12.1 Using logistic regression to analyze a dichotomous outcome variable 281
12.2 When is logistic regression used? 281
12.3 Data example 282
12.4 Basic concepts of logistic regression 283
12.5 Interpretation: categorical variables 285
12.6 Interpretation: continuous and mixed variables 288
12.7 Refining and evaluating logistic regression analysis 289
12.8 Applications of logistic regression 296
12.9 Discussion of computer programs 299
12.10 What to watch out for 301
Summary 302
References 302
Further reading 303
Problems 304

13 Regression analysis using survival data 306
13.1 Using survival analysis to analyze time-to-event data 306
13.2 When is survival analysis used? 306
13.3 Data examples 307
13.4 Survival functions 309
13.5 Common distributions used in survival analysis 314
13.6 The log-linear regression model 317
13.7 The Cox proportional hazards regression model 319
13.8 Some comparisons of the log-linear, Cox and logistic regression models 320
13.9 Discussion of computer programs 324
13.10 What to watch out for 326
Summary 327
References 328
Further reading 328
Problems 328

14 Principal components analysis 330
14.1 Using principal components analysis to understand intercorrelations 330
14.2 When is principal components analysis used? 330
14.3 Data example 331
14.4 Basic concepts of principal components analysis 333
14.5 Interpretation 336
14.6 Use of principal components analysis in regression and other applications 345
14.7 Discussion of computer programs 348
14.8 What to watch out for 350
Summary 351
References 352
Further reading 352
Problems 352

15 Factor analysis 354
15.1 Using factor analysis to examine the relationship among *P* variables 354
15.2 When is factor analysis used? 354
15.3 Data example 355
15.4 Basic concepts of factor analysis 356
15.5 Initial factor extraction: principal components analysis 358
15.6 Initial factor extraction: iterated principal components 362
15.7 Factor rotations 365
15.8 Assigning factor scores to individuals 371
15.9 An application of factor analysis to the depression data 372
15.10 Discussion of computer programs 374
15.11 What to watch out for 376
Summary 377
References 378
Further reading 379
Problems 379

16 Cluster analysis 381
16.1 Using cluster analysis to group cases 381
16.2 When is cluster analysis used? 381
16.3 Data example 383
16.4 Basic concepts: initial analysis and distance measures 385
16.5 Analytical clustering techniques 391
16.6 Cluster analysis for financial data set 398
16.7 Discussion of computer programs 404
16.8 What to watch out for 406

Summary 406
References 407
Further reading 407
Problems 408

17 Log-linear analysis 410
17.1 Using log-linear models to analyze categorical data 410
17.2 When is log-linear analysis used? 410
17.3 Data example 411
17.4 Notation and sample considerations 413
17.5 Tests of hypotheses and models for two-way tables 415
17.6 Example of a two-way table 419
17.7 Models for multiway tables 421
17.8 Tests of hypotheses for multiway tables: exploratory model building 425
17.9 Tests of hypotheses: specific models 431
17.10 Sample size issues 432
17.11 The logit model 434
17.12 Discussion of computer programs 437
17.13 What to watch out for 439
Summary 440
References 440
Further reading 441
Problems 441

Appendix A Lung function data 443
Table A.1 Code book for lung function data set 444
Table A.2 Lung function data set 445
Further reading 446

Appendix B Lung cancer survival data 446
Table B.1 Lung cancer data 447

Index 449