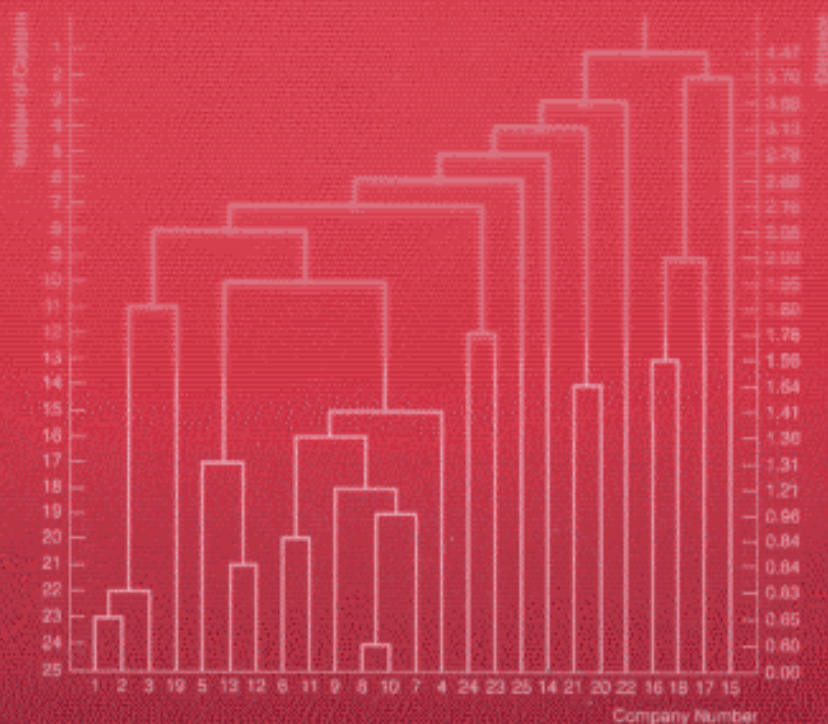


Texts in Statistical Science

Computer-Aided Multivariate Analysis

FOURTH EDITION



Abdelmonem Afifi
Virginia A. Clark and Susanne May



CHAPMAN & HALL/CRC

Contents

Preface

xiii

One	Preparation for Analysis	1
1	What is multivariate analysis?	3
1.1	Defining multivariate analysis	3
1.2	Examples of multivariate analyses	3
1.3	Multivariate analyses discussed in this book	6
1.4	Organization and content of the book	10
1.5	References	11
2	Characterizing data for analysis	13
2.1	Variables: their definition, classification, and use	13
2.2	Defining statistical variables	13
2.3	Stevens's classification of variables	14
2.4	How variables are used in data analysis	17
2.5	Examples of classifying variables	18
2.6	Other characteristics of data	19
2.7	Summary	19
2.8	References	20
2.9	Problems	20
3	Preparing for data analysis	23
3.1	Processing data so they can be analyzed	23
3.2	Choice of a statistical package	24
3.3	Techniques for data entry	26
3.4	Organizing the data	33
3.5	Example: depression study	39
3.6	Summary	43
3.7	References	43
3.8	Problems	46

4	Data screening and transformations	49
4.1	Transformations, assessing normality and independence . . .	49
4.2	Common transformations	49
4.3	Selecting appropriate transformations	53
4.4	Assessing independence	63
4.5	Summary	64
4.6	References	66
4.7	Problems	67
5	Selecting appropriate analyses	71
5.1	Which analyses to perform?	71
5.2	Why selection is often difficult	71
5.3	Appropriate statistical measures	72
5.4	Selecting appropriate multivariate analyses	76
5.5	Summary	79
5.6	References	79
5.7	Problems	80
	Two Applied Regression Analysis	83
6	Simple regression and correlation	85
6.1	Chapter outline	85
6.2	When are regression and correlation used?	86
6.3	Data example	86
6.4	Regression methods: fixed- X case	88
6.5	Regression and correlation: variable- X case	93
6.6	Interpretation: fixed- X case	94
6.7	Interpretation: variable- X case	95
6.8	Other available computer output	99
6.9	Robustness and transformations for regression	107
6.10	Other types of regression	109
6.11	Special applications of regression	113
6.12	Discussion of computer programs	116
6.13	What to watch out for	117
6.14	Summary	118
6.15	References	119
6.16	Problems	122
7	Multiple regression and correlation	125
7.1	Chapter outline	125
7.2	When are regression and correlation used?	126
7.3	Data example	126
7.4	Regression methods: fixed- X case	129
7.5	Regression and correlation: variable- X case	131

7.6	Interpretation: fixed- X case	137
7.7	Interpretation: variable- X case	140
7.8	Regression diagnostics and transformations	143
7.9	Other options in computer programs	148
7.10	Discussion of computer programs	153
7.11	What to watch out for	158
7.12	Summary	159
7.13	References	159
7.14	Problems	160
8	Variable selection in regression	165
8.1	Chapter outline	165
8.2	When are variable selection methods used?	165
8.3	Data example	166
8.4	Criteria for variable selection	170
8.5	A general F test	172
8.6	Stepwise regression	174
8.7	Subset regression	180
8.8	Discussion of computer programs	183
8.9	Discussion of strategies	186
8.10	What to watch out for	189
8.11	Summary	191
8.12	References	192
8.13	Problems	193
9	Special regression topics	197
9.1	Chapter outline	197
9.2	Missing values in regression analysis	197
9.3	Dummy variables	205
9.4	Constraints on parameters	214
9.5	Regression analysis with multicollinearity	217
9.6	Ridge regression	218
9.7	Summary	222
9.8	References	223
9.9	Problems	225
Three	Multivariate Analysis	231
10	Canonical correlation analysis	233
10.1	Chapter outline	233
10.2	When is canonical correlation analysis used?	233
10.3	Data example	234
10.4	Basic concepts of canonical correlation	235
10.5	Other topics in canonical correlation	240

10.6	Discussion of computer programs	243
10.7	What to watch out for	245
10.8	Summary	246
10.9	References	246
10.10	Problems	247
11	Discriminant analysis	249
11.1	Chapter outline	249
11.2	When is discriminant analysis used?	250
11.3	Data example	251
11.4	Basic concepts of classification	252
11.5	Theoretical background	259
11.6	Interpretation	261
11.7	Adjusting the dividing point	265
11.8	How good is the discrimination?	267
11.9	Testing variable contributions	270
11.10	Variable selection	271
11.11	Discussion of computer programs	272
11.12	What to watch out for	274
11.13	Summary	275
11.14	References	276
11.15	Problems	276
12	Logistic regression	281
12.1	Chapter outline	281
12.2	When is logistic regression used?	282
12.3	Data example	282
12.4	Basic concepts of logistic regression	284
12.5	Interpretation: Categorical variables	285
12.6	Interpretation: Continuous variables	287
12.7	Interpretation: Interactions	289
12.8	Refining and evaluating logistic regression	296
12.9	Nominal and ordinal logistic regression	308
12.10	Applications of logistic regression	315
12.11	Poisson Regression	319
12.12	Discussion of computer programs	323
12.13	What to watch out for	324
12.14	Summary	327
12.15	References	327
12.16	Problems	329

13 Regression analysis with survival data	333
13.1 Chapter outline	333
13.2 When is survival analysis used?	334
13.3 Data examples	334
13.4 Survival functions	337
13.5 Common survival distributions	343
13.6 Comparing survival among groups	344
13.7 The log-linear regression model	346
13.8 The Cox regression model	348
13.9 Comparing regression models	359
13.10 Discussion of computer programs	362
13.11 What to watch out for	364
13.12 Summary	365
13.13 References	365
13.14 Problems	367
14 Principal components analysis	369
14.1 Chapter outline	369
14.2 When is principal components analysis used?	369
14.3 Data example	370
14.4 Basic concepts	371
14.5 Interpretation	375
14.6 Other uses	383
14.7 Discussion of computer programs	386
14.8 What to watch out for	386
14.9 Summary	388
14.10 References	388
14.11 Problems	389
15 Factor analysis	391
15.1 Chapter outline	391
15.2 When is factor analysis used?	391
15.3 Data example	392
15.4 Basic concepts	393
15.5 Initial extraction: principal components	395
15.6 Initial extraction: iterated components	398
15.7 Factor rotations	402
15.8 Assigning factor scores	406
15.9 Application of factor analysis	408
15.10 Discussion of computer programs	409
15.11 What to watch out for	412
15.12 Summary	413
15.13 References	414
15.14 Problems	415

16 Cluster analysis	417
16.1 Chapter outline	417
16.2 When is cluster analysis used?	417
16.3 Data example	419
16.4 Basic concepts: initial analysis	419
16.5 Analytical clustering techniques	426
16.6 Cluster analysis for financial data set	432
16.7 Discussion of computer programs	437
16.8 What to watch out for	440
16.9 Summary	440
16.10 References	441
16.11 Problems	442
17 Log-linear analysis	445
17.1 Chapter outline	445
17.2 When is log-linear analysis used?	445
17.3 Data example	446
17.4 Notation and sample considerations	448
17.5 Tests and models for two-way tables	450
17.6 Example of a two-way table	454
17.7 Models for multiway tables	456
17.8 Exploratory model building	459
17.9 Assessing specific models	465
17.10 Sample size issues	466
17.11 The logit model	468
17.12 Discussion of computer programs	470
17.13 What to watch out for	471
17.14 Summary	473
17.15 References	474
17.16 Problems	475
Appendix A	477
A.1 Data sets and how to obtain them	477
A.2 Chemical companies financial data	477
A.3 Depression study data	477
A.4 Financial performance cluster analysis data	478
A.5 Lung cancer survival data	478
A.6 Lung function data	478
A.7 Parental HIV data	479