

Contents

Introduction	xv
Part I: Looking at multivariate data	
1 Motivation and fundamental concepts	3
1.1 Types of data	5
1.2 Towards a pictorial representation	6
1.3 Some geometrical concepts	11
1.3.1 Two-dimensional basics	11
1.3.2 Lines and subspaces	14
1.3.3 Extensions to higher dimensionality	16
1.3.4 Rotation of axes	17
1.3.5 Latent roots and vectors	22
1.4 Similarity, dissimilarity, and distance	24
2 One-way graphical representation of data matrices	33
2.1 Direct two-dimensional representations	35
2.1.1 Extensions of simple graphs	35
2.1.2 Pictorial representations	39
2.1.3 Function representations	43
2.2 Representations based on subspace projection	48
2.2.1 Geometrical models of data	48
2.2.2 Properties of the models	50
2.2.3 Principal component analysis: geometrical concepts	53
2.2.4 Principal component analysis: mathematical details	60
2.2.5 Principal component analysis: properties and practical considerations	63
2.2.6 Principal component analysis as method of data display: illustrative example	74
2.2.7 Other subspace projection methods	75
2.3 Non-numerical data and missing values	83
3 Graphical methods for association or proximity matrices	86
3.1 Direct two-dimensional representation of units, derived from numerical dissimilarities	89
3.1.1 The dendrogram	89
3.1.2 The minimum spanning tree	102

3.2	Subspace representation of units, derived from numerical dissimilarities	104
3.3	Subspace representation of units, derived from ordinal data	113
3.4	Methods for handling asymmetric matrices	120
3.5	Postscript: missing values	123
4	Two-way graphical representation of data matrices	124
4.1	The singular value decomposition of a (rectangular) matrix	126
4.2	Representing a matrix by two sets of vectors: the biplot	128
4.3	Representing a matrix by one set of points and one set of vectors: preference scaling	134
4.4	Representing a matrix by two sets of points: multi-dimensional unfolding	140
4.5	Representing incidence matrices: correspondence analysis	145
5	Analytical comparison of two or more graphical representations	153
5.1	Comparison of two n -point configurations: procrustes analysis	155
5.2	Simultaneous comparison of g n -point configurations: generalized procrustes analysis	164
5.3	Comparison of subspaces	167
Part II: Samples, populations, and models		
6	Data inspection or data analysis?	179
6.1	Basic concepts	179
6.2	Three-way scaling: individual differences analysis	183
6.3	Models for multivariate populations	193
7	Distribution theory	199
7.1	Basic concepts of multivariate distributions	199
7.2	The multivariate normal distribution	204
7.3	The Wishart distribution	209
7.4	Elliptic distributions	210
7.5	Tests for multivariate normality	211
7.6	Transformations	215
7.7	Multivariate analysis	220

Part III: Analysing ungrouped data

8	Estimation and hypothesis testing	223
8.1	Basic concepts and distributional results	223
8.2	Estimation of μ and Σ	227
8.3	Testing hypotheses about μ and Σ	235
8.4	Some comments on multivariate hypothesis tests	249
9	Reduction of dimensionality: inferential aspects of descriptive methods	252
9.1	Andrews curves	253
9.2	Principal components	254
9.3	Multidimensional scaling	259
9.4	Cluster analysis	261
9.5	Comment	264
10	Discrete data	265
10.1	Entirely discrete data: summary and model	265
10.2	Entirely discrete data: analysis	271
10.2.1	Two-dimensional tables	272
10.2.2	Three- and higher-dimensional tables	280
10.3	Mixed discrete and continuous data	282

Part IV: Analysing grouped data

11	Incorporating group structure: descriptive statistics	289
11.1	Highlighting differences between groups: canonical variates	291
11.2	Identifying common features across groups: within-group components	306
11.3	Partial information: identifying group membership	309
11.4	Miscellaneous topics	312
11.4.1	Categorical data	312
11.4.2	Missing values	316
11.4.3	Outliers and robustness	318
11.5	The need for inferential methods	321
12	Inferential aspects: the two-group case	323
12.1	Estimation and hypothesis testing	323
12.1.1	The multivariate normal model	323
12.1.2	The multinomial model	329
12.1.3	The location model	330
12.1.4	Discrimination and classification	330
12.2	Classification rules based on probability models	332

12.2.1	Fundamental principles	332
12.2.2	The multivariate normal model	339
12.2.3	The multinomial model	348
12.2.4	The location model	354
12.3	Other classification rules	356
12.3.1	Fisher's linear discriminant function	356
12.3.2	Logistic discrimination	358
12.3.3	Distance-based discrimination	359
12.4	Evaluating the performance of an allocation rule	360
12.5	A practical example	364
13	Inferential aspects: more than two groups	369
13.1	The multivariate normal model	369
13.2	Differences between the groups: canonical variates	370
13.3	Treatment structure: multivariate analysis of variance	376
13.4	Univariate repeated measurements	384
13.5	Similarities between the groups: common principal components	387
13.6	Discrimination and classification	392
13.6.1	Sample space partition	392
13.6.2	Distance-based allocation rule	397
13.6.3	Logistic discrimination	398
13.7	Non-normal data	399
Part V: Analysing association among variables		
14	Measuring and interpreting association	405
14.1	Measuring association between two variables	405
14.2	Interpreting association between two variables	414
14.3	Graphical investigation of many associations	417
14.4	Correcting correlations for effects of extraneous variables	427
14.5	Measuring association between two sets of variables	432
14.6	Testing hypotheses about sets of associations	445
14.6.1	Test that all population correlations are zero (mutual independence of all variables)	446
14.6.2	Test that all population canonical correlations are zero	447
14.6.3	Test that some population canonical correlations are zero	448
15	Exploiting observed associations: manifest-variable models	449
15.1	Motivation	449
15.2	Multivariate regression	454

15.2.1	Introduction	454
15.2.2	Fitting the model	457
15.2.3	Inference about model parameters	462
15.2.4	Assessing the adequacy of a fitted model	470
15.3	Multivariate analysis of variance revisited	472
16	Explaining observed associations: latent-variable models	474
16.1	Background ideas and principles	474
16.2	Continuous manifest variables: factor analysis	477
16.2.1	Basic model	477
16.2.2	Implications of the model	483
16.2.3	A data set	485
16.2.4	Estimation of factor loadings and specific variances	486
16.2.5	Test of goodness-of-fit, and choice of q	493
16.2.6	Factor rotation	496
16.2.7	Estimation of factor scores	500
16.2.8	Dimensionality reduction using factor analysis	502
16.2.9	Binary data	503
16.3	Discrete manifest variables: latent-structure analysis	503
16.3.1	Background	503
16.3.2	Model	505
16.3.3	Statistical aspects of the latent-class model	508
16.4	A recent unified approach	513
17	Conclusion: some general multivariate problems	516
17.1	Variable selection	516
17.2	Non-parametric assessment of error: data re-sampling schemes	520
Appendix:	some basic matrix theory	529
A1	Definitions	529
A2	Elementary arithmetic operations	531
A3	Determinants and inverses	535
A4	Quadratic forms	538
A5	Latent roots and vectors	538
A6	Matrix square root	540
A7	Partitioned matrices	541
A8	Vector differentiation	543
References		544
Index		558