

Contents

Preface	vii
Notation	xvii
1 Examples of Multivariate Data	1
1.1 Influence, Outliers and Distances	1
1.2 A Sketch of the Forward Search	3
1.3 Multivariate Normality and our Examples	5
1.4 Swiss Heads	6
1.5 National Track Records for Women	10
1.6 Municipalities in Emilia-Romagna	16
1.7 Swiss Bank Notes	22
1.8 Plan of the Book	30
2 Multivariate Data and the Forward Search	31
2.1 The Univariate Normal Distribution	32
2.1.1 Estimation	32
2.1.2 Distribution of Estimators	33
2.2 Estimation and the Multivariate Normal Distribution	34
2.2.1 The Multivariate Normal Distribution	34
2.2.2 The Wishart Distribution	35
2.2.3 Estimation of Σ	36
2.3 Hypothesis Testing	37
2.3.1 Hypotheses About the Mean	37

2.3.2	Hypotheses About the Variance	37
2.4	The Mahalanobis Distance	39
2.5	Some Deletion Results	40
2.5.1	The Deletion Mahalanobis Distance	40
2.5.2	The (Bartlett)-Sherman-Morrison-Woodbury Formula	41
2.5.3	Deletion Relationships Among Distances	42
2.6	Distribution of the Squared Mahalanobis Distance	43
2.7	Determinants of Dispersion Matrices and the Squared Mahalanobis Distance	44
2.8	Regression	46
2.9	Added Variables in Regression	49
2.10	The Mean Shift Outlier Model	51
2.11	Seemingly Unrelated Regression	53
2.12	The Forward Search	55
2.13	Starting the Search	58
2.13.1	The Babyfood Data	58
2.13.2	Robust Bivariate Boxplots from Peeling	59
2.13.3	Bivariate Boxplots from Ellipses	62
2.13.4	The Initial Subset	64
2.14	Monitoring the Search	66
2.15	The Forward Search for Regression Data	71
2.15.1	Univariate Regression	71
2.15.2	Multivariate Regression	73
2.16	Further Reading	73
2.17	Exercises	76
2.18	Solutions	78
3	Data from One Multivariate Distribution	89
3.1	Swiss Heads	89
3.2	National Track Records for Women	100
3.3	Municipalities in Emilia-Romagna	108
3.4	Swiss Bank Notes	116
3.5	What Have We Seen?	138
3.6	Exercises	140
3.7	Solutions	142
4	Multivariate Transformations to Normality	151
4.1	Background	151
4.2	An Introductory Example: the Babyfood Data	152
4.3	Power Transformations to Approximate Normality	155
4.3.1	Transformation of the Response in Regression	156
4.3.2	Multivariate Transformations to Normality	161
4.4	Score Tests for Transformations	162
4.5	Graphics for Transformations	164

4.6	Finding a Multivariate Transformation with the Forward Search	165
4.7	Babyfood Data	166
4.8	Swiss Heads	169
4.9	Horse Mussels	176
4.10	Municipalities in Emilia-Romagna	186
4.10.1	Demographic Variables	187
4.10.2	Wealth Variables	191
4.10.3	Work Variables	195
4.10.4	A Combined Analysis	200
4.11	National Track Records for Women	204
4.12	Dyestuff Data	209
4.13	Babyfood Data and Variable Selection	214
4.14	Suggestions for Further Reading	218
4.15	Exercises	220
4.16	Solutions	221
5	Principal Components Analysis	229
5.1	Background	229
5.2	Principal Components and Eigenvectors	230
5.2.1	Linear Transformations and Principal Components	230
5.2.2	Lack of Scale Invariance and Standardized Variables	232
5.2.3	The Number of Components	232
5.3	Monitoring the Forward Search	233
5.3.1	Principal Components and Variances	233
5.3.2	Principal Component Scores	234
5.3.3	Correlations Between Variables and Principal Components	235
5.3.4	Elements of the Eigenvectors	236
5.4	The Biplot and the Singular Value Decomposition	236
5.5	Swiss Heads	239
5.6	Milk Data	242
5.7	Quality of Life	252
5.8	Swiss Bank Notes	260
5.8.1	Forgeries and Genuine Notes	261
5.8.2	Forgeries Alone	263
5.9	Municipalities in Emilia-Romagna	265
5.10	Further reading	272
5.11	Exercises	274
5.12	Solutions	278
6	Discriminant Analysis	297
6.1	Background	297
6.2	An Outline of Discriminant Analysis	298
6.2.1	Bayesian Discrimination	298

6.2.2	Quadratic Discriminant Analysis	299
6.2.3	Linear Discriminant Analysis	300
6.2.4	Estimation of Means and Variances	300
6.2.5	Canonical Variates	301
6.2.6	Assessment of Discriminant Rules	304
6.3	The Forward Search	305
6.3.1	Step 1: Choice of the Initial Subset	306
6.3.2	Step 2: Adding Observations During the Forward Search	306
6.3.3	Mahalanobis Distances and Discriminant Analysis in Step 2	307
6.4	Monitoring the Search	307
6.5	Transformations to Normality in Discriminant Analysis . .	309
6.6	Iris Data	310
6.7	Electrodes Data	317
6.8	Transformed Iris Data	324
6.9	Swiss Bank Notes	328
6.10	Importance of Transformations in Discriminant Analysis: A Simulated Example	332
6.10.1	A Deletion Analysis	332
6.10.2	Finding a Transformation with the Forward Search .	337
6.10.3	Discriminant Analysis and Confirmation of the Transformation	341
6.11	Muscular Dystrophy Data	344
6.11.1	The Data	344
6.11.2	Finding the Transformation	345
6.11.3	Outliers and Discriminant Analysis	349
6.11.4	More Data	351
6.12	Further reading	356
6.13	Exercises	357
6.14	Solutions	359
7	Cluster Analysis	367
7.1	Introduction	367
7.2	Clustering and the Forward Search	368
7.2.1	Three Steps in Finding Clusters	368
7.2.2	Standardized Mahalanobis Distances and Analysis with Many Clusters	369
7.2.3	Forward Searches in Cluster Analysis	370
7.3	The 60:80 Data	371
7.3.1	Failure of a Very Robust Statistical Method	372
7.3.2	The Forward Search	373
7.3.3	Further Plots for the 60:80 Data	375
7.4	Three Clusters, Two Outliers: A Second Synthetic Example	379
7.4.1	A Forward Analysis	379

7.4.2	A Very Robust Analysis	382
7.5	Data with a Bridge	385
7.5.1	Preliminary Analysis	386
7.5.2	Further Preliminary Analysis: Mahalanobis Distances for Groups and Individual Units	392
7.5.3	Exploratory Analysis: Single Clusters for the Bridge Data	398
7.5.4	Confirmatory Analysis: Three Clusters for the Bridge Data	401
7.6	Financial Data	406
7.6.1	Preliminary Analysis	406
7.6.2	Exploratory Analysis	410
7.6.3	Confirmatory Analysis	417
7.7	Diabetes Data	420
7.7.1	Preliminary Analysis	420
7.7.2	Exploratory Analysis	428
7.7.3	Confirmatory Analysis	436
7.8	Discussion	439
7.8.1	Agglomerative Hierarchical Clustering	441
7.8.2	Partitioning Methods	443
7.8.3	Some Examples from Traditional Cluster Analysis	444
7.8.4	Model-Based Clustering	446
7.8.5	Further Reading	448
7.9	Exercises	450
7.10	Solutions	451
8	Spatial Linear Models	457
8.1	Introduction	457
8.2	Background on Kriging	459
8.2.1	Ordinary Kriging	459
8.2.2	Isotropic Semivariogram Models	465
8.2.3	Spatial Outliers	467
8.2.4	Kriging Diagnostics	468
8.2.5	Robust Estimation of the Variogram	471
8.3	The Forward Search for Ordinary Kriging	472
8.3.1	Choice of the Initial Subset	472
8.3.2	Progressing in the Search	474
8.3.3	Monitoring the Search	475
8.4	Contaminated Kriging Examples	477
8.4.1	Multiple Spatial Outliers	477
8.4.2	Pocket of Nonstationarity	479
8.5	Wheat Yield Data	482
8.6	Reflectance Data	491
8.7	Background on Spatial Autoregression	495
8.7.1	Neighbourhood Structure and Edge Correction	498

8.7.2	Simultaneous Spatial Autoregression (SAR) Models	501
8.7.3	Spatial Outliers Under the SAR Model	502
8.7.4	High Leverage Sites	504
8.8	The Block Forward Search for Spatial Autoregression	506
8.8.1	Subset Likelihood	508
8.8.2	Defining the Blocks	509
8.8.3	Choice of the Initial Subset	510
8.8.4	Progressing in the Search	511
8.8.5	Monitoring the Search	511
8.9	SAR Examples With Multiple Contamination	513
8.9.1	Masked Spatial Outliers	513
8.9.2	Estimation of ρ	516
8.9.3	Multiple High Leverage Sites	519
8.10	Wheat Yield Data Revisited	522
8.11	Further Reading	524
8.12	Exercises	526
8.13	Solutions	528
	Appendix: Tables of Data	551
	Bibliography	597
	Author Index	607
	Subject Index	611

Exploring Multivariate Data with the Forward Search

Atkinson, A.; Riani, M.; cccc, A.

2004, XXIV, 624 p., Hardcover

ISBN: 978-0-387-40852-1