

Wolfgang Karl Härdle · Léopold Simar

Applied Multivariate Statistical Analysis

Fifth Edition

 Springer

Contents

Part I Descriptive Techniques

1	Comparison of Batches	3
1.1	Boxplots	4
1.2	Histograms	11
1.3	Kernel Densities	13
1.4	Scatterplots	18
1.5	Chernoff-Flury Faces	21
1.6	Andrews' Curves	23
1.7	Parallel Coordinate Plots	27
1.8	Hexagon Plots	32
1.9	Boston Housing	35
1.10	Exercises	42
	References	43

Part II Multivariate Random Variables

2	A Short Excursion into Matrix Algebra	47
2.1	Elementary Operations	47
2.2	Spectral Decompositions	53
2.3	Quadratic Forms	55
2.4	Derivatives	58
2.5	Partitioned Matrices	59
2.6	Geometrical Aspects	61
2.7	Exercises	68
3	Moving to Higher Dimensions	71
3.1	Covariance	72
3.2	Correlation	76

3.3	Summary Statistics	81
3.4	Linear Model for Two Variables	84
3.5	Simple Analysis of Variance	91
3.6	Multiple Linear Model	95
3.7	Boston Housing	100
3.8	Exercises	103
	References	105
4	Multivariate Distributions	107
4.1	Distribution and Density Function	108
4.2	Moments and Characteristic Functions	113
4.3	Transformations	123
4.4	The Multinormal Distribution	125
4.5	Sampling Distributions and Limit Theorems	129
4.6	Heavy-Tailed Distributions	135
4.7	Copulae	151
4.8	Bootstrap	161
4.9	Exercises	164
	References	166
5	Theory of the Multinormal	167
5.1	Elementary Properties of the Multinormal	167
5.2	The Wishart Distribution	174
5.3	Hotelling's T^2 -Distribution	176
5.4	Spherical and Elliptical Distributions	178
5.5	Exercises	180
	References	182
6	Theory of Estimation	183
6.1	The Likelihood Function	184
6.2	The Cramer–Rao Lower Bound	188
6.3	Exercises	192
	Reference	193
7	Hypothesis Testing	195
7.1	Likelihood Ratio Test	196
7.2	Linear Hypothesis	205
7.3	Boston Housing	222
7.4	Exercises	225
	References	229

Part III Multivariate Techniques

8 Regression Models 233

8.1 General ANOVA and ANCOVA Models 235

8.1.1 ANOVA Models 235

8.1.2 ANCOVA Models 240

8.1.3 Boston Housing 242

8.2 Categorical Responses 243

8.2.1 Multinomial Sampling and Contingency Tables 243

8.2.2 Log-Linear Models for Contingency Tables 244

8.2.3 Testing Issues with Count Data 248

8.2.4 Logit Models 251

8.3 Exercises 258

Reference 259

9 Variable Selection 261

9.1 Lasso 262

9.1.1 Lasso in the Linear Regression Model 262

9.1.2 Lasso in High Dimensions 271

9.1.3 Lasso in Logit Model 272

9.2 Elastic Net 276

9.2.1 Elastic Net in Linear Regression Model 277

9.2.2 Elastic Net in Logit Model 278

9.3 Group Lasso 279

9.4 Exercises 282

References 283

10 Decomposition of Data Matrices by Factors 285

10.1 The Geometric Point of View 286

10.2 Fitting the p -Dimensional Point Cloud 287

10.3 Fitting the n -Dimensional Point Cloud 290

10.4 Relations Between Subspaces 292

10.5 Practical Computation 293

10.6 Exercises 296

11 Principal Components Analysis 299

11.1 Standardized Linear Combination 300

11.2 Principal Components in Practice 303

11.3 Interpretation of the PCs 307

11.4 Asymptotic Properties of the PCs 310

11.5 Normalized Principal Components Analysis 313

11.6 Principal Components as a Factorial Method 315

11.7 Common Principal Components 320

11.8	Boston Housing	323
11.9	More Examples	326
11.10	Exercises	335
	References	336
12	Factor Analysis	337
12.1	The Orthogonal Factor Model	338
12.2	Estimation of the Factor Model	345
12.3	Factor Scores and Strategies	352
12.4	Boston Housing	355
12.5	Exercises	358
	References	361
13	Cluster Analysis	363
13.1	The Problem	364
13.2	The Proximity Between Objects	365
13.3	Cluster Algorithms	370
13.4	Adaptive Weights Clustering	381
13.5	Spectral Clustering	385
13.6	Boston Housing	388
13.7	Exercises	391
	References	393
14	Discriminant Analysis	395
14.1	Allocation Rules for Known Distributions	395
14.2	Discrimination Rules in Practice	402
14.3	Boston Housing	408
14.4	Exercises	410
	References	411
15	Correspondence Analysis	413
15.1	Motivation	413
15.2	Chi-Square Decomposition	416
15.3	Correspondence Analysis in Practice	420
15.4	Exercises	429
16	Canonical Correlation Analysis	431
16.1	Most Interesting Linear Combination	431
16.2	Canonical Correlation in Practice	436
16.3	Exercises	441
	References	442
17	Multidimensional Scaling	443
17.1	The Problem	443
17.2	Metric Multidimensional Scaling	447

17.3	Nonmetric Multidimensional Scaling	452
17.4	Exercises	458
	References	459
18	Conjoint Measurement Analysis	461
18.1	Introduction	461
18.2	Design of Data Generation	463
18.3	Estimation of Preference Orderings	465
18.4	Exercises	472
	References	473
19	Applications in Finance	475
19.1	Portfolio Choice	475
19.2	Efficient Portfolio	476
19.3	Efficient Portfolios in Practice	483
19.4	The Capital Asset Pricing Model (CAPM)	484
19.5	Exercises	486
	Reference	486
20	Computationally Intensive Techniques	487
20.1	Simplicial Depth	488
20.2	Projection Pursuit	491
20.3	Sliced Inverse Regression	496
20.4	Support Vector Machines	504
20.5	Classification and Regression Trees	519
20.6	Boston Housing	534
20.7	Exercises	538
	References	539
 Part IV Appendix		
21	Symbols and Notations	543
21.1	Basics	543
21.2	Mathematical Abbreviations	543
21.3	Samples	544
21.4	Densities and Distribution Functions	544
21.5	Moments	544
21.6	Empirical Moments	545
21.7	Distributions	545
22	Data	547
22.1	Boston Housing Data	547
22.2	Swiss Bank Notes	547
22.3	Car Data	548

22.4	Classic Blue Pullovers Data	548
22.5	U.S. Companies Data	549
22.6	French Food Data	549
22.7	Car Marks	549
22.8	U.S. Crime Data	549
22.9	Bankruptcy Data I	550
22.10	Bankruptcy Data II	551
22.11	Journaux Data	551
22.12	Timebudget Data	551
22.13	Vocabulary Data	552
22.14	French Baccalauréat Frequencies	553
	References	553
	Index	555