Contents

Preface to the 2nd Edition v

Preface to the 1st Edition vii

I Descriptive Techniques 1

1 Comparison of Batches 3
1.1 Boxplots . . . 4
1.2 Histograms . . . 10
1.3 Kernel Densities . . . 13
1.4 Scatterplots . . . 17
1.5 Chernoff-Flury Faces . . . 20
1.6 Andrews' Curves . . . 23
1.7 Parallel Coordinate Plots . . . 27
1.8 Boston Housing . . . 28
1.9 Exercises . . . 35

II Multivariate Random Variables 39

2 A Short Excursion into Matrix Algebra 41
2.1 Elementary Operations . . . 41
2.2 Spectral Decompositions . . . 46
2.3 Quadratic Forms . . . 47
2.4 Derivatives . . . 50
2.5 Partitioned Matrices . . . 51
2.6 Geometrical Aspects . . . 52
2.7 Exercises . . . 59

3 Moving to Higher Dimensions 61
3.1 Covariance . . . 61
3.2 Correlation . . . 65
3.3 Summary Statistics . . . 70
3.4 Linear Model for Two Variables . . . 73
3.5 Simple Analysis of Variance . . . 79
3.6 Multiple Linear Model . . . 83
3.7 Boston Housing . . . 87
3.8 Exercises . . . 90

4 Multivariate Distributions 93
4.1 Distribution and Density Function . . . 93
4.2 Moments and Characteristic Functions . . . 98
4.3 Transformations . . . 106
4.4 The Multinormal Distribution . . . 108
4.5 Sampling Distributions and Limit Theorems . . . 111
4.6 Heavy-Tailed Distributions . . . 118
4.7 Copulae . . . 132
4.8 Bootstrap . . . 141
4.9 Exercises . . . 144

5 Theory of the Multinormal 147
5.1 Elementary Properties of the Multinormal . . . 147
5.2 The Wishart Distribution . . . 153
5.3 Hotelling's T^2-Distribution . . . 154
5.4 Spherical and Elliptical Distributions . . . 156
5.5 Exercises . . . 158

6 Theory of Estimation 161
6.1 The Likelihood Function . . . 161
6.2 The Cramer-Rao Lower Bound . . . 165
6.3 Exercises . . . 168

7 Hypothesis Testing 171
7.1 Likelihood Ratio Test . . . 171
7.2 Linear Hypothesis . . . 179
7.3 Boston Housing . . . 194
7.4 Exercises . . . 196

III Multivariate Techniques 201

8 Decomposition of Data Matrices by Factors 203
8.1 The Geometric Point of View . . . 203
8.2 Fitting the p-dimensional Point Cloud . . . 205
8.3 Fitting the n-dimensional Point Cloud . . . 208
8.4 Relations between Subspaces . . . 209
8.5 Practical Computation . . . 211
8.6 Exercises . . . 213

9 Principal Components Analysis 215
9.1 Standardized Linear Combination . . . 215
9.2 Principal Components in Practice . . . 219
9.3 Interpretation of the PCs . . . 222
9.4 Asymptotic Properties of the PCs . . . 226
9.5 Normalized Principal Components Analysis . . . 228
9.6 Principal Components as a Factorial Method . . . 229
9.7 Common Principal Components . . . 234
9.8 Boston Housing . . . 237
9.9 More Examples . . . 239
9.10 Exercises . . . 247

10 Factor Analysis **251**
10.1 The Orthogonal Factor Model . . . 251
10.2 Estimation of the Factor Model . . . 257
10.3 Factor Scores and Strategies . . . 264
10.4 Boston Housing . . . 265
10.5 Exercises . . . 269

11 Cluster Analysis **271**
11.1 The Problem . . . 271
11.2 The Proximity between Objects . . . 272
11.3 Cluster Algorithms . . . 276
11.4 Boston Housing . . . 284
11.5 Exercises . . . 285

12 Discriminant Analysis **289**
12.1 Allocation Rules for Known Distributions . . . 289
12.2 Discrimination Rules in Practice . . . 295
12.3 Boston Housing . . . 300
12.4 Exercises . . . 301

13 Correspondence Analysis **305**
13.1 Motivation . . . 305
13.2 Chi-square Decomposition . . . 307
13.3 Correspondence Analysis in Practice . . . 310
13.4 Exercises . . . 318

14 Canonical Correlation Analysis **321**
14.1 Most Interesting Linear Combination . . . 321
14.2 Canonical Correlation in Practice . . . 325
14.3 Exercises . . . 330

15 Multidimensional Scaling **331**
15.1 The Problem . . . 331
15.2 Metric Multidimensional Scaling . . . 336
15.3 Nonmetric Multidimensional Scaling . . . 339
15.4 Exercises . . . 346

16 Conjoint Measurement Analysis **347**
16.1 Introduction . . . 347
16.2 Design of Data Generation . . . 349
16.3 Estimation of Preference Orderings . . . 351
16.4 Exercises . . . 356

17 Applications in Finance **359**
17.1 Portfolio Choice . . . 359
17.2 Efficient Portfolio . . . 360
17.3 Efficient Portfolios in Practice . . . 365
17.4 The Capital Pricing Model (CAPM) . . . 367
17.5 Exercises . . . 368

18 Computationally Intensive Techniques **371**
18.1 Simplicial Depth . . . 371
18.2 Projection Pursuit . . . 375

18.3 Sliced Inverse Regression . . . 379
18.4 Support Vector Machines . . . 385
18.5 Classification and Regression Trees . . . 401
18.6 Boston Housing . . . 417
18.7 Exercises . . . 418

IV Appendix 421

A Symbols and Notations 423

B Data 427
B.1 Boston Housing Data . . . 427
B.2 Swiss Bank Notes . . . 427
B.3 Car Data . . . 430
B.4 Classic Blue Pullovers Data . . . 432
B.5 U.S. Companies Data . . . 432
B.6 French Food Data . . . 434
B.7 Car Marks . . . 434
B.8 French Baccalauréat Frequencies . . . 435
B.9 Journaux Data . . . 435
B.10 U.S. Crime Data . . . 436
B.11 Plasma Data . . . 437
B.12 WAIS Data . . . 438
B.13 ANOVA Data . . . 439
B.14 Timebudget Data . . . 440
B.15 Geopol Data . . . 441
B.16 U.S. Health Data . . . 443
B.17 Vocabulary Data . . . 444
B.18 Athletic Records Data . . . 445
B.19 Unemployment Data . . . 447
B.20 Annual Population Data . . . 447
B.21 Bankruptcy Data . . . 448

Bibliography 451

Index 455