# Contents


**I Descriptive Techniques** 11

**1 Comparison of Batches** 13

1.1 Boxplots . . . . . 14

1.2 Histograms . . . . . 22

1.3 Kernel Densities . . . . . 25

1.4 Scatterplots . . . . . 30

1.5 Chernoff-Flury Faces . . . . . 34

1.6 Andrews' Curves . . . . . 39

1.7 Parallel Coordinates Plots . . . . . 42

1.8 Boston Housing . . . . . 44

1.9 Exercises . . . . . 52

**II Multivariate Random Variables** 55

**2 A Short Excursion into Matrix Algebra** 57

2.1 Elementary Operations . . . . . 57

2.2 Spectral Decompositions . . . . . 63

2.3 Quadratic Forms . . . . . 65

2.4 Derivatives . . . . . 68

2.5 Partitioned Matrices . . . . . 68

2.6 Geometrical Aspects . . . . . 71
2.7 Exercises . . . . . 79

**3 Moving to Higher Dimensions** **81**

3.1 Covariance . . . . . 82
3.2 Correlation . . . . . 86
3.3 Summary Statistics . . . . . 92
3.4 Linear Model for Two Variables . . . . . 95
3.5 Simple Analysis of Variance . . . . . 103
3.6 Multiple Linear Model . . . . . 108
3.7 Boston Housing . . . . . 112
3.8 Exercises . . . . . 115

**4 Multivariate Distributions** **119**

4.1 Distribution and Density Function . . . . . 120
4.2 Moments and Characteristic Functions . . . . . 125
4.3 Transformations . . . . . 135
4.4 The Multinormal Distribution . . . . . 137
4.5 Sampling Distributions and Limit Theorems . . . . . 142
4.6 Bootstrap . . . . . 148
4.7 Exercises . . . . . 152

**5 Theory of the Multinormal** **155**

5.1 Elementary Properties of the Multinormal . . . . . 155
5.2 The Wishart Distribution . . . . . 162
5.3 Hotelling Distribution . . . . . 165
5.4 Spherical and Elliptical Distributions . . . . . 167
5.5 Exercises . . . . . 169

**6 Theory of Estimation** **173**

6.1 The Likelihood Function . . . . . 174

6.2 The Cramer-Rao Lower Bound . . . . . 178

6.3 Exercises . . . . . 181

**7 Hypothesis Testing** **183**

7.1 Likelihood Ratio Test . . . . . 184

7.2 Linear Hypothesis . . . . . 192

7.3 Boston Housing . . . . . 209

7.4 Exercises . . . . . 212

**III Multivariate Techniques** **217**

**8 Decomposition of Data Matrices by Factors** **219**

8.1 The Geometric Point of View . . . . . 220

8.2 Fitting the $p$-dimensional Point Cloud . . . . . 221

8.3 Fitting the $n$-dimensional Point Cloud . . . . . 225

8.4 Relations between Subspaces . . . . . 227

8.5 Practical Computation . . . . . 228

8.6 Exercises . . . . . 232

**9 Principal Components Analysis** **233**

9.1 Standardized Linear Combinations . . . . . 234

9.2 Principal Components in Practice . . . . . 238

9.3 Interpretation of the PCs . . . . . 241

9.4 Asymptotic Properties of the PCs . . . . . 246

9.5 Normalized Principal Components Analysis . . . . . 249

9.6 Principal Components as a Factorial Method . . . . . 250

9.7 Common Principal Components . . . . . 256

9.8 Boston Housing . . . . . 259
9.9 More Examples . . . . . 261
9.10 Exercises . . . . . 272

**10 Factor Analysis** **275**

10.1 The Orthogonal Factor Model . . . . . 275
10.2 Estimation of the Factor Model . . . . . 282
10.3 Factor Scores and Strategies . . . . . 291
10.4 Boston Housing . . . . . 293
10.5 Exercises . . . . . 298

**11 Cluster Analysis** **301**

11.1 The Problem . . . . . 301
11.2 The Proximity between Objects . . . . . 302
11.3 Cluster Algorithms . . . . . 308
11.4 Boston Housing . . . . . 316
11.5 Exercises . . . . . 318

**12 Discriminant Analysis** **323**

12.1 Allocation Rules for Known Distributions . . . . . 323
12.2 Discrimination Rules in Practice . . . . . 331
12.3 Boston Housing . . . . . 337
12.4 Exercises . . . . . 339

**13 Correspondence Analysis** **341**

13.1 Motivation . . . . . 341
13.2 Chi-square Decomposition . . . . . 344
13.3 Correspondence Analysis in Practice . . . . . 347
13.4 Exercises . . . . . 358

**14 Canonical Correlation Analysis** **361**

14.1 Most Interesting Linear Combination . . . . . 361

14.2 Canonical Correlation in Practice . . . . . 366

14.3 Exercises . . . . . 372

**15 Multidimensional Scaling** **373**

15.1 The Problem . . . . . 373

15.2 Metric Multidimensional Scaling . . . . . 379

15.2.1 The Classical Solution . . . . . 379

15.3 Nonmetric Multidimensional Scaling . . . . . 383

15.3.1 Shepard-Kruskal algorithm . . . . . 384

15.4 Exercises . . . . . 391

**16 Conjoint Measurement Analysis** **393**

16.1 Introduction . . . . . 393

16.2 Design of Data Generation . . . . . 395

16.3 Estimation of Preference Orderings . . . . . 398

16.4 Exercises . . . . . 405

**17 Applications in Finance** **407**

17.1 Portfolio Choice . . . . . 407

17.2 Efficient Portfolio . . . . . 408

17.3 Efficient Portfolios in Practice . . . . . 415

17.4 The Capital Asset Pricing Model (CAPM) . . . . . 417

17.5 Exercises . . . . . 418

**18 Highly Interactive, Computationally Intensive Techniques** **421**

18.1 Simplicial Depth . . . . . 421

18.2 Projection Pursuit . . . . . 425

18.3 Sliced Inverse Regression . . . . . 431

18.4 Boston Housing . . . . . 439
18.5 Exercises . . . . . 440

**A Symbols and Notation** **443**

**B Data** **447**

B.1 Boston Housing Data . . . . . 447
B.2 Swiss Bank Notes . . . . . 448
B.3 Car Data . . . . . 452
B.4 Classic Blue Pullovers Data . . . . . 454
B.5 U.S. Companies Data . . . . . 455
B.6 French Food Data . . . . . 457
B.7 Car Marks . . . . . 458
B.8 French Baccalauréat Frequencies . . . . . 459
B.9 Journaux Data . . . . . 460
B.10 U.S. Crime Data . . . . . 461
B.11 Plasma Data . . . . . 463
B.12 WAIS Data . . . . . 464
B.13 ANOVA Data . . . . . 466
B.14 Timebudget Data . . . . . 467
B.15 Geopol Data . . . . . 469
B.16 U.S. Health Data . . . . . 471
B.17 Vocabulary Data . . . . . 473
B.18 Athletic Records Data . . . . . 475
B.19 Unemployment Data . . . . . 477
B.20 Annual Population Data . . . . . 478

**Bibliography** **479**

**Index** **483**