Contents

Preface *page* xiii

1 Basic Concepts in Probability and Statistics 1
- 1.1 Graphical Description of Data 2
- 1.2 Measures of Central Value: Mean, Median, and Mode 4
- 1.3 Measures of Variation: Percentile Ranges and Variance 6
- 1.4 Population versus a Sample 8
- 1.5 Elements of Probability Theory 8
- 1.6 Expectation 11
- 1.7 More Than One Random Variable 13
- 1.8 Independence 16
- 1.9 Estimating Population Quantities from Samples 18
- 1.10 Normal Distribution and Associated Theorems 20
- 1.11 Independence versus Zero Correlation 27
- 1.12 Further Topics 28
- 1.13 Conceptual Questions 29

2 Hypothesis Tests 30
- 2.1 The Problem 31
- 2.2 Introduction to Hypothesis Testing 33
- 2.3 Further Comments on the t-test 40
- 2.4 Examples of Hypothesis Tests 43
- 2.5 Summary of Common Significance Tests 49
- 2.6 Further Topics 50
- 2.7 Conceptual Questions 51

3 Confidence Intervals 52
- 3.1 The Problem 53

3.2 Confidence Interval for a Difference in Means 53
3.3 Interpretation of the Confidence Interval 55
3.4 A Pitfall about Confidence Intervals 57
3.5 Common Procedures for Confidence Intervals 57
3.6 Bootstrap Confidence Intervals 64
3.7 Further Topics 67
3.8 Conceptual Questions 68

4 Statistical Tests Based on Ranks 69
4.1 The Problem 70
4.2 Exchangeability and Ranks 71
4.3 The Wilcoxon Rank-Sum Test 73
4.4 Stochastic Dominance 78
4.5 Comparison with the *t*-test 79
4.6 Kruskal–Wallis Test 81
4.7 Test for Equality of Dispersions 83
4.8 Rank Correlation 85
4.9 Derivation of the Mean and Variance of the Rank Sum 88
4.10 Further Topics 92
4.11 Conceptual Questions 93

5 Introduction to Stochastic Processes 94
5.1 The Problem 95
5.2 Stochastic Processes 100
5.3 Why Should I Care if My Data Are Serially Correlated? 105
5.4 The First-Order Autoregressive Model 109
5.5 The AR(2) Model 117
5.6 Pitfalls in Interpreting ACFs 119
5.7 Solutions of the AR(2) Model 121
5.8 Further Topics 122
5.9 Conceptual Questions 124

6 The Power Spectrum 126
6.1 The Problem 127
6.2 The Discrete Fourier Transform 129
6.3 Parseval's Identity 133
6.4 The Periodogram 134
6.5 The Power Spectrum 135
6.6 Periodogram of Gaussian White Noise 138
6.7 Impact of a Deterministic Periodic Component 139

6.8 Estimation of the Power Spectrum 140
6.9 Presence of Trends and Jump Discontinuities 144
6.10 Linear Filters 146
6.11 Tying Up Loose Ends 150
6.12 Further Topics 152
6.13 Conceptual Questions 155

7 Introduction to Multivariate Methods 156
7.1 The Problem 157
7.2 Vectors 159
7.3 The Linear Transformation 160
7.4 Linear Independence 163
7.5 Matrix Operations 166
7.6 Invertible Transformations 168
7.7 Orthogonal Transformations 170
7.8 Random Vectors 172
7.9 Diagonalizing a Covariance Matrix 175
7.10 Multivariate Normal Distribution 178
7.11 Hotelling's T-squared Test 179
7.12 Multivariate Acceptance and Rejection Regions 181
7.13 Further Topics 182
7.14 Conceptual Questions 183

8 Linear Regression: Least Squares Estimation 185
8.1 The Problem 186
8.2 Method of Least Squares 188
8.3 Properties of the Least Squares Solution 192
8.4 Geometric Interpretation of Least Squares Solutions 196
8.5 Illustration Using Atmospheric CO_2 Concentration 199
8.6 The Line Fit 205
8.7 Always Include the Intercept Term 206
8.8 Further Topics 207
8.9 Conceptual Questions 209

9 Linear Regression: Inference 210
9.1 The Problem 211
9.2 The Model 212
9.3 Distribution of the Residuals 212
9.4 Distribution of the Least Squares Estimates 213
9.5 Inferences about Individual Regression Parameters 215

9.6 Controlling for the Influence of Other Variables 216
9.7 Equivalence to "Regressing Out" Predictors 218
9.8 Seasonality as a Confounding Variable 222
9.9 Equivalence between the Correlation Test and Slope Test 224
9.10 Generalized Least Squares 225
9.11 Detection and Attribution of Climate Change 226
9.12 The General Linear Hypothesis 233
9.13 Tying Up Loose Ends 234
9.14 Conceptual Questions 236

10 Model Selection 237
10.1 The Problem 238
10.2 Bias–Variance Trade off 240
10.3 Out-of-Sample Errors 243
10.4 Model Selection Criteria 245
10.5 Pitfalls 249
10.6 Further Topics 253
10.7 Conceptual Questions 254

11 Screening: A Pitfall in Statistics 255
11.1 The Problem 256
11.2 Screening *iid* Test Statistics 259
11.3 The Bonferroni Procedure 262
11.4 Screening Based on Correlation Maps 262
11.5 Can You Trust Relations Inferred from Correlation Maps? 265
11.6 Screening Based on Change Points 265
11.7 Screening with a Validation Sample 268
11.8 The Screening Game: Can You Find the Statistical Flaw? 268
11.9 Screening Always Exists in Some Form 271
11.10 Conceptual Questions 272

12 Principal Component Analysis 273
12.1 The Problem 274
12.2 Examples 276
12.3 Solution by Singular Value Decomposition 283
12.4 Relation between PCA and the Population 285
12.5 Special Considerations for Climate Data 289
12.6 Further Topics 295
12.7 Conceptual Questions 297

13 Field Significance 298
13.1 The Problem 299
13.2 The Livezey–Chen Field Significance Test 303
13.3 Field Significance Test Based on Linear Regression 305
13.4 False Discovery Rate 310
13.5 Why Different Tests for Field Significance? 311
13.6 Further Topics 312
13.7 Conceptual Questions 312

14 Multivariate Linear Regression 314
14.1 The Problem 315
14.2 Review of Univariate Regression 317
14.3 Estimating Multivariate Regression Models 320
14.4 Hypothesis Testing in Multivariate Regression 323
14.5 Selecting X 324
14.6 Selecting Both X and Y 328
14.7 Some Details about Regression with Principal Components 331
14.8 Regression Maps and Projecting Data 332
14.9 Conceptual Questions 333

15 Canonical Correlation Analysis 335
15.1 The Problem 336
15.2 Summary and Illustration of Canonical Correlation Analysis 337
15.3 Population Canonical Correlation Analysis 343
15.4 Relation between CCA and Linear Regression 347
15.5 Invariance to Affine Transformation 349
15.6 Solving CCA Using the Singular Value Decomposition 350
15.7 Model Selection 357
15.8 Hypothesis Testing 359
15.9 Proof of the Maximization Properties 362
15.10 Further Topics 364
15.11 Conceptual Questions 364

16 Covariance Discriminant Analysis 366
16.1 The Problem 367
16.2 Illustration: Most Detectable Climate Change Signals 370
16.3 Hypothesis Testing 378
16.4 The Solution 382
16.5 Solution in a Reduced-Dimensional Subspace 388
16.6 Variable Selection 392

16.7 Further Topics 395
16.8 Conceptual Questions 398

17 Analysis of Variance and Predictability 399
17.1 The Problem 400
17.2 Framing the Problem 401
17.3 Test Equality of Variance 403
17.4 Test Equality of Means: ANOVA 404
17.5 Comments about ANOVA 406
17.6 Weather Predictability 407
17.7 Measures of Predictability 411
17.8 What Is the Difference between Predictability and Skill? 414
17.9 Chaos and Predictability 416
17.10 Conceptual Questions 417

18 Predictable Component Analysis 418
18.1 The Problem 419
18.2 Illustration of Predictable Component Analysis 422
18.3 Multivariate Analysis of Variance 424
18.4 Predictable Component Analysis 427
18.5 Variable Selection in PrCA 430
18.6 PrCA Based on Other Measures of Predictability 432
18.7 Skill Component Analysis 435
18.8 Connection to Multivariate Linear Regression and CCA 437
18.9 Further Properties of PrCA 439
18.10 Conceptual Questions 445

19 Extreme Value Theory 446
19.1 The Problem and a Summary of the Solution 447
19.2 Distribution of the Maximal Value 453
19.3 Maximum Likelihood Estimation 459
19.4 Nonstationarity: Changing Characteristics of Extremes 463
19.5 Further Topics 466
19.6 Conceptual Questions 467

20 Data Assimilation 468
20.1 The Problem 469
20.2 A Univariate Example 469
20.3 Some Important Properties and Interpretations 473
20.4 Multivariate Gaussian Data Assimilation 475

20.5 Sequential Processing of Observations 477
20.6 Multivariate Example 478
20.7 Further Topics 481
20.8 Conceptual Questions 487

21 Ensemble Square Root Filters 489
21.1 The Problem 490
21.2 Filter Divergence 497
21.3 Monitoring the Innovations 499
21.4 Multiplicative Inflation 500
21.5 Covariance Localization 503
21.6 Further Topics 507
21.7 Conceptual Questions 509

Appendix 510
A.1 Useful Mathematical Relations 510
A.2 Generalized Eigenvalue Problems 511
A.3 Derivatives of Quadratic Forms and Traces 512

References 514

Index 523