

Contents

Preface to the Second Edition	xv
Preface to the First Edition	xvii
Symbols and Abbreviations	xix
1 Introduction	1
1.1 Deterministic Data and Random Data	1
1.2 Population, Sample and Statistics	5
1.3 Random Variables	8
1.4 Probabilities and Distributions	10
1.4.1 Discrete Variables	10
1.4.2 Continuous Variables	12
1.5 Beyond a Reasonable Doubt	13
1.6 Statistical Significance and Other Significances	17
1.7 Datasets	19
1.8 Software Tools	19
1.8.1 SPSS and STATISTICA	20
1.8.2 MATLAB and R	22
2 Presenting and Summarising the Data	29
2.1 Preliminaries	29
2.1.1 Reading in the Data	29
2.1.2 Operating with the Data	34
2.2 Presenting the Data	39
2.2.1 Counts and Bar Graphs	40
2.2.2 Frequencies and Histograms	47
2.2.3 Multivariate Tables, Scatter Plots and 3D Plots	52
2.2.4 Categorical Plots	56
2.3 Summarising the Data	58
2.3.1 Measures of Location	58
2.3.2 Measures of Spread	62
2.3.3 Measures of Shape	64

2.3.4	Measures of Association for Continuous Variables.....	66
2.3.5	Measures of Association for Ordinal Variables.....	69
2.3.6	Measures of Association for Nominal Variables.....	73
	Exercises.....	77
3	Estimating Data Parameters	81
3.1	Point Estimation and Interval Estimation.....	81
3.2	Estimating a Mean	85
3.3	Estimating a Proportion	92
3.4	Estimating a Variance	95
3.5	Estimating a Variance Ratio.....	97
3.6	Bootstrap Estimation.....	99
	Exercises.....	107
4	Parametric Tests of Hypotheses	111
4.1	Hypothesis Test Procedure.....	111
4.2	Test Errors and Test Power	115
4.3	Inference on One Population.....	121
4.3.1	Testing a Mean	121
4.3.2	Testing a Variance	125
4.4	Inference on Two Populations	126
4.4.1	Testing a Correlation	126
4.4.2	Comparing Two Variances.....	129
4.4.3	Comparing Two Means	132
4.5	Inference on More than Two Populations.....	141
4.5.1	Introduction to the Analysis of Variance.....	141
4.5.2	One-Way ANOVA	143
4.5.3	Two-Way ANOVA	156
	Exercises.....	166
5	Non-Parametric Tests of Hypotheses	171
5.1	Inference on One Population.....	172
5.1.1	The Runs Test.....	172
5.1.2	The Binomial Test.....	174
5.1.3	The Chi-Square Goodness of Fit Test	179
5.1.4	The Kolmogorov-Smirnov Goodness of Fit Test.....	183
5.1.5	The Lilliefors Test for Normality	187
5.1.6	The Shapiro-Wilk Test for Normality	187
5.2	Contingency Tables.....	189
5.2.1	The 2×2 Contingency Table	189
5.2.2	The $r \times c$ Contingency Table	193

5.2.3	The Chi-Square Test of Independence	195
5.2.4	Measures of Association Revisited.....	197
5.3	Inference on Two Populations	200
5.3.1	Tests for Two Independent Samples.....	201
5.3.2	Tests for Two Paired Samples	205
5.4	Inference on More Than Two Populations.....	212
5.4.1	The Kruskal-Wallis Test for Independent Samples.....	212
5.4.2	The Friedmann Test for Paired Samples	215
5.4.3	The Cochran Q test.....	217
	Exercises.....	218

6 Statistical Classification 223

6.1	Decision Regions and Functions	223
6.2	Linear Discriminants.....	225
6.2.1	Minimum Euclidian Distance Discriminant	225
6.2.2	Minimum Mahalanobis Distance Discriminant.....	228
6.3	Bayesian Classification	234
6.3.1	Bayes Rule for Minimum Risk.....	234
6.3.2	Normal Bayesian Classification	240
6.3.3	Dimensionality Ratio and Error Estimation.....	243
6.4	The ROC Curve	246
6.5	Feature Selection.....	253
6.6	Classifier Evaluation	256
6.7	Tree Classifiers	259
	Exercises.....	268

7 Data Regression 271

7.1	Simple Linear Regression	272
7.1.1	Simple Linear Regression Model	272
7.1.2	Estimating the Regression Function	273
7.1.3	Inferences in Regression Analysis.....	279
7.1.4	ANOVA Tests	285
7.2	Multiple Regression	289
7.2.1	General Linear Regression Model	289
7.2.2	General Linear Regression in Matrix Terms	289
7.2.3	Multiple Correlation	292
7.2.4	Inferences on Regression Parameters.....	294
7.2.5	ANOVA and Extra Sums of Squares.....	296
7.2.6	Polynomial Regression and Other Models	300
7.3	Building and Evaluating the Regression Model.....	303
7.3.1	Building the Model.....	303
7.3.2	Evaluating the Model	306
7.3.3	Case Study	308
7.4	Regression Through the Origin.....	314

7.5	Ridge Regression	316
7.6	Logit and Probit Models	322
	Exercises.....	327
8	Data Structure Analysis	329
8.1	Principal Components.....	329
8.2	Dimensional Reduction.....	337
8.3	Principal Components of Correlation Matrices.....	339
8.4	Factor Analysis	347
	Exercises.....	350
9	Survival Analysis	353
9.1	Survivor Function and Hazard Function	353
9.2	Non-Parametric Analysis of Survival Data	354
	9.2.1 The Life Table Analysis	354
	9.2.2 The Kaplan-Meier Analysis.....	359
	9.2.3 Statistics for Non-Parametric Analysis.....	362
9.3	Comparing Two Groups of Survival Data	364
9.4	Models for Survival Data	367
	9.4.1 The Exponential Model	367
	9.4.2 The Weibull Model.....	369
	9.4.3 The Cox Regression Model	371
	Exercises.....	373
10	Directional Data	375
10.1	Representing Directional Data	375
10.2	Descriptive Statistics	380
10.3	The von Mises Distributions	383
10.4	Assessing the Distribution of Directional Data.....	387
	10.4.1 Graphical Assessment of Uniformity	387
	10.4.2 The Rayleigh Test of Uniformity	389
	10.4.3 The Watson Goodness of Fit Test	392
	10.4.4 Assessing the von Misesness of Spherical Distributions.....	393
10.5	Tests on von Mises Distributions	395
	10.5.1 One-Sample Mean Test	395
	10.5.2 Mean Test for Two Independent Samples	396
10.6	Non-Parametric Tests.....	397
	10.6.1 The Uniform Scores Test for Circular Data.....	397
	10.6.2 The Watson Test for Spherical Data.....	398
	10.6.3 Testing Two Paired Samples	399
	Exercises.....	400

Appendix A - Short Survey on Probability Theory 403

- A.1 Basic Notions403
 - A.1.1 Events and Frequencies403
 - A.1.2 Probability Axioms.....404
- A.2 Conditional Probability and Independence406
 - A.2.1 Conditional Probability and Intersection Rule.....406
 - A.2.2 Independent Events406
- A.3 Compound Experiments.....408
- A.4 Bayes' Theorem.....409
- A.5 Random Variables and Distributions410
 - A.5.1 Definition of Random Variable410
 - A.5.2 Distribution and Density Functions411
 - A.5.3 Transformation of a Random Variable413
- A.6 Expectation, Variance and Moments414
 - A.6.1 Definitions and Properties414
 - A.6.2 Moment-Generating Function417
 - A.6.3 Chebyshev Theorem.....418
- A.7 The Binomial and Normal Distributions.....418
 - A.7.1 The Binomial Distribution418
 - A.7.2 The Laws of Large Numbers419
 - A.7.3 The Normal Distribution420
- A.8 Multivariate Distributions422
 - A.8.1 Definitions.....422
 - A.8.2 Moments.....425
 - A.8.3 Conditional Densities and Independence.....425
 - A.8.4 Sums of Random Variables427
 - A.8.5 Central Limit Theorem.....428

Appendix B - Distributions 431

- B.1 Discrete Distributions431
 - B.1.1 Bernoulli Distribution.....431
 - B.1.2 Uniform Distribution432
 - B.1.3 Geometric Distribution.....433
 - B.1.4 Hypergeometric Distribution434
 - B.1.5 Binomial Distribution435
 - B.1.6 Multinomial Distribution436
 - B.1.7 Poisson Distribution438
- B.2 Continuous Distributions439
 - B.2.1 Uniform Distribution439
 - B.2.2 Normal Distribution.....441
 - B.2.3 Exponential Distribution.....442
 - B.2.4 Weibull Distribution444
 - B.2.5 Gamma Distribution445
 - B.2.6 Beta Distribution446
 - B.2.7 Chi-Square Distribution.....448

B.2.8	Student's t Distribution	449
B.2.9	F Distribution	451
B.2.10	Von Mises Distributions	452
Appendix C - Point Estimation		455
C.1	Definitions.....	455
C.2	Estimation of Mean and Variance	457
Appendix D - Tables		459
D.1	Binomial Distribution	459
D.2	Normal Distribution	465
D.3	Student's t Distribution	466
D.4	Chi-Square Distribution	467
D.5	Critical Values for the F Distribution	468
Appendix E - Datasets		469
E.1	Breast Tissue	469
E.2	Car Sale	469
E.3	Cells	470
E.4	Clays	470
E.5	Cork Stoppers.....	471
E.6	CTG	472
E.7	Culture	473
E.8	Fatigue	473
E.9	FHR.....	474
E.10	FHR-Apgar	474
E.11	Firms	475
E.12	Flow Rate	475
E.13	Foetal Weight.....	475
E.14	Forest Fires.....	476
E.15	Freshmen.....	476
E.16	Heart Valve	477
E.17	Infarct.....	478
E.18	Joints	478
E.19	Metal Firms.....	479
E.20	Meteo	479
E.21	Moulds	479
E.22	Neonatal	480
E.23	Programming.....	480
E.24	Rocks	481
E.25	Signal & Noise	481

E.26	Soil Pollution	482
E.27	Stars	482
E.28	Stock Exchange.....	483
E.29	VCG	484
E.30	Wave	484
E.31	Weather	484
E.32	Wines	485

Appendix F - Tools **487**

F.1	MATLAB Functions.....	487
F.2	R Functions	488
F.3	Tools EXCEL File	489
F.4	SCSize Program.....	489

References **491**

Index **499**