Contents

PART I BASIC PRINCIPLES 1

CHAPTER 1 INTRODUCTION 3

1.1 *Human error and ignorance* 10
1.2 *Outliers in relation to probability models* 11
1.3 *Outliers in more structured situations* 16
1.4 *Bayesian methods* 23
1.5 *Statistical computing* 24
1.6 *Survey of outlier problems* 24

CHAPTER 2 WHY DO OUTLYING OBSERVATIONS ARISE AND WHAT SHOULD ONE DO ABOUT THEM? 27

2.1 *Early informal approaches* 27
2.2 *Origin of outliers, statistical methods and aims* 31
2.2.1 *The nature and origin of an outlier* 32
2.2.2 *Relevant statistical procedures for handling outliers* 34
2.2.3 *Different aims in examining outliers* 39
2.3 *Models for contamination* 43
2.4 *Outlier proneness, outlier resistance, outlier labelling* 52

CHAPTER 3 THE ACCOMMODATION APPROACH: ROBUST ESTIMATION AND TESTING 55

3.1 *Performance criteria* 60
3.1.1 *Efficiency measures for estimators* 60
3.1.2 *Distributional properties of contaminated distributions* 66
3.1.3 *The qualitative approach: influence curves* 68
3.1.4 *Robustness of confidence intervals* 74
3.1.5 *Robustness of significance tests* 75
3.2 *Methods of accommodation—general and outlier-specific* 78
3.2.1 *Estimation of location* 78
3.2.2 *Estimation of dispersion* 83
3.2.3 *Hypothesis tests and confidence intervals* 85
3.2.4 *Adaptive robust procedures* 87

CHAPTER 4 TESTING FOR DISCORDANCY: PRINCIPLES AND CRITERIA 89

4.1 *Construction of discordancy tests* 94
4.1.1 *Test statistics* 94
4.1.2 *Statistical bases for construction of tests* 98
4.1.3 *Inclusive and exclusive measures* 108
4.1.4 *Masking and swamping* 109
4.1.5 *Assessment of significance* 115
4.2 *The assessment of test performance* 121
4.2.1 *Measures of performance* 121
4.2.2 *Distributional properties under the alternative hypothesis* 125
4.3 *The multiple outlier problem* 125
4.3.1 *Block procedures for multiple outliers in univariate samples* 133
4.3.2 *Consecutive procedures for multiple outliers in univariate samples* 136

PART II UNIVARIATE DATA 141

CHAPTER 5 ACCOMMODATION PROCEDURES FOR UNIVARIATE SAMPLES 143

5.1 *Estimation of location* 143
5.1.1 *Estimators based on trimming or Winsorization* 143
5.1.2 *L-estimators (linear order statistics estimators)* 146
5.1.3 *M-estimators (maximum likelihood type estimators)* 148
5.1.4 *R-estimators (rank test estimators)* 152
5.1.5 *Other estimators* 153
5.2 *Estimation of scale or dispersion* 155
5.3 *Hypothesis tests and confidence intervals* 157
5.4 *Accommodation of outliers in univariate normal samples* 158
5.5 *Accommodation of outliers in gamma (including exponential) samples* 174
5.6 *Accommodation of outliers in logistic and double exponential samples* 185
5.6.1 *Estimators of the mean* μ 186
5.6.2 *Estimators of the scale parameter* σ 187
5.6.3 *Robust estimation for logistic samples* 188
5.6.4 *Robust estimation for double exponential samples* 189

CHAPTER 6 SPECIFIC DISCORDANCY TESTS FOR OUTLIERS IN UNIVARIATE SAMPLES 191

6.1 *Guide to use of the tests* 191
6.2 *Discordancy tests for gamma (including exponential) samples* 193
6.2.1 *Gamma samples: contents list and details of tests* 195
6.3 *Discordancy tests for normal samples* 216
6.3.1 *Normal samples: contents list and details of tests* 217
6.4 *Discordancy tests for samples from other distributions* 250
6.4.1 *Log-normal samples* 251
6.4.2 *Truncated exponential samples* 251
6.4.3 *Uniform samples* 252
6.4.4 *Gumbel, Fréchet, and Weibull samples* 254
6.4.5 *Pareto samples* 260
6.4.6 *Poisson samples* 261
6.4.7 *Binomial samples* 264

PART III MULTIVARIATE AND STRUCTURED DATA 267

CHAPTER 7 OUTLIERS IN MULTIVARIATE DATA 269

7.1 *Principles for outlier detection in multivariate samples* 269
7.2 *Accommodation of multivariate outliers* 273
7.2.1 *Estimation of parameters in multivariate distributions* 273
7.2.2 *Estimation of individual components of vector parameters* 279
7.2.3 *Outliers in multivariate analyses* 282
7.3 *Discordancy tests* 283
7.3.1 *Multivariate normal samples* 284
7.3.2 *Multivariate exponential samples* 293
7.3.3 *Multivariate Pareto samples* 295
7.3.4 *Discordancy tests for more general distributions* 296
7.3.5 *A transformation approach* 296
7.4 *Informal methods for multivariate outliers* 297
7.4.1 *Marginal outliers and linear constraints* 300
7.4.2 *Graphical and pictorial methods* 301
7.4.3 *Principal component analysis method* 303
7.4.4 *Use of reduction measures in the form of generalized distances* 306
7.4.5 *Function plots* 308
7.4.6 *Correlation methods and influential observations* 309
7.4.7 *A 'gap test' for multivariate outliers* 311

CHAPTER 8 THE OUTLIER PROBLEM FOR STRUCTURED DATA: REGRESSION, THE LINEAR MODEL AND DESIGNED EXPERIMENTS 315

8.1 *Outliers in simple linear regression* 320
8.1.1 *Discordancy tests for simple linear regression* 321
8.1.2 *Outlier accommodation in simple linear regression* 323
8.1.3 *Outliers in linear structural and functional models* 325
8.2 *Outliers with general linear models* 328
8.2.1 *Residual-based methods* 328
8.2.2 *Augmented residual-based, and non-residual-based, methods* 340
8.2.3 *Accommodation of outliers for the linear model* 346
8.3 *Outliers in non-linear regression* 351
8.4 *Outliers in designed experiments* 354
8.5 *Graphical methods and diagnostics for linear model outliers* 365
8.6 *Some further comments on influential observations* 371

PART IV SPECIAL TOPICS 375

CHAPTER 9 BAYESIAN APPROACHES TO OUTLIERS 377

9.1 *Basic Bayesian considerations* 377
9.2 *Bayesian accommodation of outliers* 380
9.3 *Bayesian assessment of contamination* 388

CHAPTER 10 OUTLIERS IN TIME SERIES: AN IMPORTANT AREA OF OUTLIER STUDY 395

10.1 *Detection and testing of outliers in time series* 396
10.2 *Accommodation of outliers in time series* 401
10.2.1 *Time-domain characteristics* 401
10.2.2 *Frequency-domain characteristics* 407
10.2.3 *Specific inference problems* 409
10.3 *Bayesian methods* 413
10.4 *Comment* 414

CHAPTER 11 OUTLIERS IN DIRECTIONAL DATA 417

11.1 *Outliers on the circle* 420
11.2 *Outliers on the sphere* 424
11.2.1 *Vectorial data* 424
11.2.2 *Axial data* 428

CHAPTER 12 SOME LITTLE-EXPLORED AREAS: CONTINGENCY TABLES AND SAMPLE SURVEYS 431

12.1 *Outliers in contingency tables* 431
12.2 *Outliers in sample surveys* 440

CHAPTER 13 IMPORTANT STRANDS: COMPUTER SOFTWARE, DATA STUDIES, STANDARDS AND REGULATIONS 449

13.1 *Outliers and statistical computing* 449
13.2 *Outliers in detailed data studies* 454
13.3 *Outliers in standards requirements* 455

CHAPTER 14 PERSPECTIVE 459

APPENDIX: STATISTICAL TABLES 463

Contents list 465

REFERENCES AND BIBLIOGRAPHY 527

INDEX 574