

Contents

<i>Preface</i>	vii
<i>Acknowledgements</i>	xi
<i>Abbreviations</i>	xiii

Part I: BASIC

1. Introduction	3
1.1 Statistics and Biostatistics	3
1.1.1 Types of Data	4
1.1.2 Variables and their Types	4
1.2 History and Applications	7
1.2.1 History	7
1.2.2 Applications	9
2. Data Handling I: Graphical Methods	10
2.1 Classification and Tabulation	10
2.1.1 Classification	10
2.1.2 Frequency Tables	11
2.2 Graphical Methods	14
2.2.1 Graphical Methods for Qualitative Data	15
2.2.2 Graphical Methods for Quantitative Data	20
3. Data Handling II: Descriptive Statistics	28
3.1 Measures of Location or Measures of Central Tendency	29
3.2 Measures of Dispersion	40
3.3 Measures of Skewness and Kurtosis	51
3.3.1 Moments	52
3.3.2 Sheppard Correction for Moments	54
3.3.3 Measures of Skewness and Kurtosis	54
3.4 Absolute Measures of Dispersion	55
4. Concepts of Probability	57
4.1 Uncertainty and Random Experiments	57
4.2 Sample Space and Events	59
4.2.1 Definition of Sample Space	59
4.2.2 Definition of Events	61
4.2.3 Types of Events (Definitions)	61

4.3	Definitions of Probability	64
4.3.1	Classical Definition	64
4.3.2	Statistical Definition	66
4.3.3	Axiomatic Definition of Probability	67
4.4	Additive Rule of Probability	67
4.5	Multiplicative Rule	69
4.5.1	Conditional Probability	70
4.5.2	Independent Events	71
4.6	The Bayes Rule or Bayes Theorem	73
5.	Random Variables and their Characteristics	75
5.1	Definition and Types of Random Variables	75
5.1.1	Definition of Random Variable	75
5.1.2	Types of Random Variables	76
5.2	Functions for Probability Distribution of a Random Variable	77
5.2.1	Probability Mass Function (pmf)	77
5.2.2	Probability Density Function (pdf)	78
5.2.3	Probability Distribution of Random Variable	80
5.2.4	Cumulative Distribution Function (cdf)	81
5.3	Joint pmf, Joint pdf, Marginal and Conditional pdf and Independent Random Variables	83
5.3.1	Joint pmf and Joint pdf	84
5.3.2	Marginal and Conditional Distributions	86
5.3.3	Independent Random Variables	89
5.4	Expected Values of Random Variables and their Rules	90
5.4.1	Rules for the Expected Values	91
5.5	Expected Values of Function of Random Variables	92
5.6	Generating Functions	93
5.6.1	Probability Generating Function	93
5.6.2	Moment Generating Function	93
5.6.3	Characteristic Function	94
5.7	Raw and Central Moments	95
5.7.1	Raw Moments	95
5.7.2	Central Moments	97
5.8	Coefficients of Skewness and Kurtosis	97
6.	Distributions	98
6.1	Discrete and Continuous Distributions	98
6.2	Binomial Distribution	98
6.2.1	Properties of Binomial Distribution	100
6.3	Poisson Distribution	101
6.3.1	Properties of Poisson Distribution	104
6.4	Hypergeometric Distribution	105
6.4.1	Properties of Hypergeometric Distribution	105
6.5	Geometric Distribution	106
6.5.1	Properties of Geometric Distribution	107
6.6	Negative Binomial Distribution	107
6.6.1	Properties of Negative Binomial Distribution	108

6.7	Normal Distribution	109
6.7.1	Properties of Normal Distribution	110
6.8	Uniform and Rectangular Distributions	113
6.8.1	Properties of Rectangular Distribution	114
6.9	Bivariate Normal Distribution	115
6.10	Chi-square Distribution	118
6.10.1	Properties of Chi-square Distribution	119
6.11	Student's <i>t</i> -Distribution	120
6.11.1	Properties of <i>t</i> -Distribution	121
6.12	<i>F</i> -Distribution	121
6.12.1	Properties of <i>F</i> -Distribution	122
7.	Biostatistical Inference	124
7.1	Inference	124
7.1.1	Examples of Use of Inductive Inference	125
7.2	General Concepts	125
7.3	Estimation	129
7.3.1	Point and Interval Estimation	129
7.3.2	Criteria for a Good Estimator	131
7.3.3	Methods of Estimation	132
7.4	Testing of Hypothesis	136
7.4.1	Two Types of Errors	137
7.4.2	Procedure of Testing of Hypothesis	138
8.	Tests of Significance	140
8.1	One Sample Problems for Testing Mean	142
8.2	Two Sample Problems for Testing Means	148
8.3	One Sample Problems for Testing Variance	156
8.4	Two Sample Problems for Testing Variances	160
8.5	Comparing Several Variances: Bartlett's Test	162
8.6	Comparison of Several Means	165
9.	Bivariate and Multivariate Data: Measuring and Testing Relationship	170
9.1	Simple or Pearson's Product Moment Correlation Coefficient	170
9.2	Simple Linear Regression	177
9.3	Tests of Correlation	184
9.4	Tests of Regression Coefficient	187
9.5	Testing Homogeneity of Correlation and Regression Coefficients	191
9.6	Intraclass and Spearman's Rank Correlation	200
10.	Analysis of Categorical Data: Independence and Association	206
10.1	Two Categories: Estimation and Tests of Proportions	206
10.2	Testing Independence and Homogeneity in 2×2 and $r \times c$ Contingency Table	210
11.	Electronic Data Handling	221
11.1	Introduction to Computers	221
11.2	Man-Machine Communication: Binary Code and High Level Languages	226
11.3	Working on DOS, Windows, MS Office and Computer Networks	230

Part II: ADVANCED

12. Types and Architecture of Studies	239
12.1 Planning of Experiments in Lab and in Fields	239
12.2 Design of Experiments (DoE)	242
12.3 Case-control, Cross Sectional, Longitudinal Studies and Clinical Trials	245
12.3.1 Observational Cohort Studies and Longitudinal Studies	247
12.3.2 Clinical Trials	251
12.3.3 Case-control Studies	253
12.3.4 Cross-sectional Studies	257
12.3.5 Advantages and Disadvantages of Various Studies	258
13. Data Collection: Census and Sampling	260
13.1 Census of Human Population and Animal Population	260
13.2 Random Sampling from Theoretical Distribution and from Finite Population	267
13.2.1 Selection of Random Sample from a Theoretical Distribution	267
13.2.2 Random Sampling from a Finite Population	271
13.3 Stratified Random Sampling	276
13.4 Cluster Sampling and Area Sampling	279
13.5 Systematic Sampling	280
13.6 Two-stage and Multistage Sampling	283
13.7 Purposive or Judgement Sampling	285
13.8 Snowball Sampling	285
13.9 Probability Sampling	286
14. Analysis of Data: With Violated Assumptions and from Complex Designs	287
14.1 Comparison of Two Means when Variances are Unequal	287
14.1.1 Comparison of Several Means and Completely Randomised Design	291
14.2 Randomised Block Design	293
14.3 Latin Square Design (LS_qD)	302
14.4 Factorial Analysis	307
14.4.1 2^2 Factorial Experiment	308
14.4.2 $p \times q$ Factorial Experiment	313
14.5 Nested Designs	317
14.6 BIBD and PBIBD	319
14.6.1 Balanced Incomplete Block Design (BIBD)	322
14.6.2 Partially Balanced Incomplete Block Design (PBIBD)	323
14.7 Multiple Comparisons	325
14.7.1 Equal Number of Replications or Equal Sample Sizes	325
14.7.2 Unequal Number of Replications or Unequal Sample Sizes	332
14.7.3 Multiple Comparison in Two Factor ANOVA	333
15. Non-Parametric Methods I: One Sample Tests	334
15.1 χ^2 -Test of Goodness of Fit	335
15.2 Kolmogorov-Smirnov Test	342
15.3 Sign Test	345
15.4 Wilcoxon Signed Rank Test	346

16. Non-Parametric Methods II: Two Sample Tests	349
16.1 Sign Test for Two Samples	349
16.2 Median Test	351
16.3 Wald-Wolfowitz Runs Test	354
16.4 Wilcoxon Signed Rank Test	356
16.5 Wilcoxon-Mann-Whitney <i>U</i> -Test	358
16.6 Kolmogorov-Smirnov Two Sample Test	359
17. Non-Parametric Methods III: <i>k</i>-Sample Tests	362
17.1 Median Test for <i>k</i> -Samples	362
17.2 Kruskal-Wallis <i>k</i> -Sample Test	365
17.3 Friedman's Test for RBD	367
17.4 Median Test for Two-Way Classification	369
17.5 Olmstead-Tukey Corner (or Quadrant Sum) Test of Association	371
17.6 Coefficient of Concordance and Kendall's Tau Coefficient	373
18. Time Series Analysis	377
18.1 Components of Time Series and their Determination	380
18.1.1 Determination of Components of Time Series	381
18.2 Autocorrelation in Time Series	395
18.3 Stationarity in Time Series, Transformation and Tests of Stationarity	399
18.3.1 Tests of Stationarity in Time Series	400
18.3.2 Transformation of Non-Stationary Time Series	401
18.4 Prediction or Forecasting	402
19. Bioassay	403
19.1 Types of Biological Assays, Direct Assays	403
19.1.1 Direct Assays	406
19.1.2 Dilution Assays	410
19.2 Indirect Assays and Dose Response Relationship	414
19.2.1 The Dose Response Regression	414
19.2.2 Methods of Estimation of Potency	418
19.2.3 Parallel Line Assay	420
19.2.4 Slope Ratio Assay	425
19.3 Quantal Response Assays	426
19.3.1 Probit Analysis	427
19.3.2 Logit Analysis	429
19.4 Estimation of Potency	430
19.5 Computational Procedure by Probit Analysis	431
20. Multivariate Analysis I	435
20.1 Hotelling's T^2 and Mahalanobis D^2	445
20.2 Discriminant Analysis: Classification in Two or More than Two Populations	448
20.3 MANOVA	452
21. Multivariate Analysis II	455
21.1 Principal Component Analysis (PCA)	460
21.2 Factor Analysis	464

21.2.1 Mathematical Formulation of Factor Analysis Model	465
21.2.2 Factor Analysis Procedures	468
21.2.3 Test of Number of Factors	471
21.2.4 Interpretation of Factors	472
21.2.5 Factor Rotation	473
21.2.6 Factor Scores	473
21.3 Cluster Analysis	481
21.3.1 Distance and Similarity Matrices	482
21.3.2 Clustering Methods	485
22. Bioinformatics and Computational Biology	497
22.1 Concepts of Bioinformatics: A Digital Laboratory	497
22.2 Databases and Tools of Bioinformatics	498
22.3 Sequence Analysis	502
22.3.1 Protein Sequences	510
22.3.2 FASTA and BLAST	513
22.4 Application of Hidden Markov Model (HMM)	514
22.5 Microarray Data	521
22.6 Probabilistic Modelling and Clustering of Microarray Data	523
22.6.1 Statistical Significance of Search (or Alignment)	524
22.6.2 Cluster Analysis of Microarray Data	525
23. Computer Techniques	527
23.1 Programming in FORTRAN and C++	527
23.1.1 Programming in FORTRAN	528
23.1.2 Programming in C and C++	542
23.2 Use of Statistical Packages	551
23.2.1 SPSS	552
23.2.2 BMDP	553
23.2.3 SAS	554

APPENDICES

<i>Appendix A: Statistical and Mathematical Tables</i>	559
<i>Appendix B: Mathematical Symbols and Expressions</i>	654
<i>Appendix C: Basics of Matrix Algebra</i>	657
<i>Appendix D: Elements of Set Theory</i>	675
<i>References</i>	677
<i>Subject Index</i>	689
<i>Author Index</i>	693