



MEDICAL STATISTICS and COMPUTER EXPERIMENTS

2nd Edition

Editor

Ji-Qian Fang

Sun Yat-Sen University, P R China

with

Yongyong Xu

*Fourth Military Medical
University, P R China*

Songlin Yu

*Huazhong University of Science
and Technology, P R China*

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Contents

Preface to the Second Edition	v
Introduction	ix
About the Editors	xxi
Part I Basic Concepts	1
Chapter 1. Descriptive Statistics	3
1.1 Variables and Data	3
1.2 Frequency Table and Histogram	6
1.3 Measurement for Average Level of a Sample	13
1.4 Measurement for Variation of a Sample	20
1.5 Relative Measures and Standardization Approaches	22
1.6 Frequently Used Graphs in Statistics	28
1.7 Computerized Experiments	35
1.8 Practice and Experiments	41
Chapter 2. Probability and Distribution	45
2.1 Explanation of Probability and Related Concepts	45
2.2 Distributional Characters of Random Variables	49
2.3 Binomial Distribution	53
2.4 Poisson Distribution	60
2.5 Normal Distribution	63
2.6 Computerized Experiments	71
2.7 Practice and Experiments	74

Chapter 3. Sampling Error and Confidence Interval	77
3.1 The Distribution of Sample Mean	77
3.2 t Distribution	83
3.3 The Confidence Interval for Population Mean of a Normal Distribution	85
3.4 Four Confidence Intervals for Probability and the Difference between Two Probabilities	87
3.5 The Sample Size for Estimation of Confidence Interval	88
3.6 Computerized Experiments	90
3.7 Practice and Experiments	93
Chapter 4. Hypothesis Testing for Continuous Variables	95
4.1 Specific Logic and Main Steps of Hypothesis Testing	95
4.2 The t Test for One Group of Data under Completely Randomized Design	99
4.3 The t Test for Data under Randomized Paired Design	101
4.4 The Tests for Comparing Two Means Based on Two Groups of Data under Completely Randomized Design	103
4.5 The F -Test for Equal Variances of Two Groups of Data under Completely Randomized Design	107
4.6 Test for Normality	110
4.7 The Z -Test for the Parameters of Binomial Distribution and Poisson Distribution (Large Sample)	112
4.8 Computerized Experiments	121
4.9 Practice and Experiments	125
Chapter 5. Chi-Square Test for Categorical Variable	131
5.1 Chi-Square Distribution and Pearson's Goodness-of-Fit Test	131
5.2 The χ^2 Test for Comparison between Two Independent Sample Proportions	134
5.3 The χ^2 Tests for Binary Variable under a Paired Design	142
5.4 The χ^2 Test for $R \times C$ Contingency Table	148
5.5 The χ^2 Test for Confirming a Supposed Distribution	153
5.6 Hypothesis Testing for Two Standardized Rates	155
5.7 Fisher's Exact Test for 2×2 Table	159

5.8	Computerized Experiments	163
5.9	Practice and Experiments	166
Chapter 6. Further Discussion on Hypothesis Test		171
6.1	Two Types of Error and Power	172
6.2	The Four Elements Affecting the Power	174
6.3	The Quantitative Relation between Power and the Four Elements	177
6.4	Estimation of Sample Size for the Tests in Common Use . . .	182
6.5	Non-Inferiority Test and Equivalence Test	185
6.6	Permutation Test	189
6.7	Computerized Experiments	192
6.8	Practice and Experiments	194
Chapter 7. Single-Factor Analysis of Variance		197
7.1	One-Way Analysis of Variance: Completely Random Design	197
7.2	Two-Way Analysis of Variance: Randomized Complete-Block Design	215
7.3	Three-Way Analysis of Variance: The Latin-Square Design	221
7.4	Computerized Experiments	229
7.5	Practice and Experiments	233
Chapter 8. Nonparametric Test Based on Ranks		237
8.1	Wilcoxon's Signed Rank Test	238
8.2	Wilcoxon's Rank-Sum Test for Comparing the Locations of Two Distributions	242
8.3	Hypothesis Testing for the Locations of More Than Two Populations	248
8.4	Computerized Experiments	257
8.5	Practice and Experiments	259
Chapter 9. Simple Linear Correlation		263
9.1	Concept of Correlation	263
9.2	Correlation Coefficient	266
9.3	Inference on Correlation Coefficient	269

9.4	Rank Correlation	272
9.5	Caution in Analysis of Linear Correlation	275
9.6	Computerized Experiments	277
9.7	Practice and Experiments	278

Chapter 10. Simple Linear Regression **281**

10.1	Statistical Description of Linear Regression	281
10.2	Statistical Inference on Regression	284
10.3	Applications of Linear Regression and the Pre-requisites	292
10.4	On the Basic Assumptions and Analysis of Residuals	299
10.5	Non-linear Regression	301
10.6	Computerized Experiments	309
10.7	Practice and Experiments	313

Chapter 11. Statistical Principles for Design of Interventional Study **317**

11.1	The Essential Concepts of Design	318
11.2	Statistical Principle in Clinical Trials	323
11.3	Randomization Techniques	332
11.4	Randomized Controlled Trial	336
11.5	Comments on Some Medical Examples	342
11.6	Computerized Experiments	345
11.7	Practice and Experiments	347

Part II Multi-variate Statistics **349**

Chapter 12. Multiple Regression and Correlation **351**

12.1	Basic Procedure of Multiple Regression	351
12.2	Multiple Correlation	357
12.3	Selection of Independent Variables	361
12.4	Further Topics in Multiple Regression	365
12.5	Path Analysis	373
12.6	Computerized Experiments	378
12.7	Practice and Experiments	380

Chapter 13. Measures of Multi-variate Data and Multi-variate Analysis of Variance	383
13.1 Multi-variate Statistical Description	383
13.2 Comparison between Two Mean Vectors — Hotelling's T^2 Test	388
13.3 Comparisons among Several Multi-variate Means—Multi-variate Analysis of Variance	392
13.4 Computerized Experiments	397
13.5 Practice and Experiments	400
Chapter 14. Discriminant Analysis	403
14.1 Basic Ideas of Discriminant Analysis	403
14.2 Fisher's Discriminant Analysis	405
14.3 Bayesian Discriminant Analysis	407
14.4 Stepwise Discriminant Function	411
14.5 Decision Tree	413
14.6 Retrospective and Prospective Validation	422
14.7 Considerations in Applications	424
14.8 Computerized Experiments	426
14.9 Practice and Experiments	428
Chapter 15. Logistic Regression	431
15.1 Logistic Regression Model	431
15.2 Conditional Logistic Regression	445
15.3 Multinomial Logistic Regression Model	447
15.4 Two-Level Logistic Mixed Effects Regression Model	453
15.5 Application of Logistic Regression	456
15.6 Computerized Experiments	458
15.7 Practice and Experiments	462
Chapter 16. Cluster Analysis	465
16.1 The Meaning of Clustering	465
16.2 Hierarchical Cluster	467
16.3 Fast Cluster	471
16.4 Variable Cluster	473

16.5	Computerized Experiments	474
16.6	Practice and Experiments	477
Chapter 17. Principal Component Analysis		479
17.1	The Basic Concepts of Principal Component Analysis	479
17.2	Computation and Interpretation of Principal Components	483
17.3	Principal Component Analysis in Regression	487
17.4	Computerized Experiments	490
17.5	Practice and Experiments	493
Chapter 18. Factor Analysis		497
18.1	Factor Model	497
18.2	Derivation of Factors	498
18.3	Factor Pattern Plot and Factor Rotation	502
18.4	Factor Score and Application of Factor Patterns	507
18.5	Confirmatory Factor Analysis	508
18.6	Computerized Experiments	514
18.7	Practice and Experiments	515
Chapter 19. Canonical Correlation and Correspondence Analysis		517
19.1	Canonical Correlation	517
19.2	Correspondence Analysis	528
19.3	Canonical Discriminant Analysis	535
19.4	Computerized Experiments	538
19.5	Practice and Experiments	539
Chapter 20. Survival Analysis		541
20.1	The Basic Concept of Survival Analysis	542
20.2	The Product-Limit Method for One Group of Survival Data	543
20.3	The Log-Rank Test and Breslow Test for Comparing Two Survival Data Sets	547
20.4	The Cox Regression	552
20.5	Computerized Experiments	558
20.6	Practice and Experiments	558

Chapter 21. Log-Linear Model for Contingency Table and Poisson Regression	561
21.1 Log-Linear Models for Contingency Table	561
21.2 Poisson Regression	574
21.3 Computerized Experiments	578
21.4 Practice and Experiments	581
Part III Design and Analysis for Medical Research	583
Chapter 22. Multi-Factor Analysis of Variance	585
22.1 Factorial Experiments and Analysis of Variance	585
22.2 Split-Plot Designs and Analysis of Variance	593
22.3 Cross-Over Design and Analysis of Variance	604
22.4 Computerized Experiments	609
22.5 Practice and Experiments	611
Chapter 23. Analysis of Repeated Continuous-Type Measurements	615
23.1 Examples of Repeated Measurements	615
23.2 Imperfect Analysis and its Origins	618
23.3 Approach with Summary Measures	619
23.4 Analysis of Variance for Repeated Measurements	620
23.5 Computerized Experiments	629
23.6 Practice and Experiments	631
Chapter 24. Design and Analysis of Cross-Sectional Studies	633
24.1 Design of the Study	633
24.2 Sampling Methods and Estimation of Population Parameters	634
24.3 Estimation of Sample Size	642
24.4 The Current Life Table	648
24.5 Computerized Experiments	657
24.6 Practice and Experiments	659
Chapter 25. Design and Analysis of Prospective Studies	661
25.1 Study Design	661
25.2 Measures of Disease Occurrence	663

25.3	Analysis of Data from Prospective Studies	671
25.4	Computerized Experiments	684
25.5	Practice and Experiments	692
Chapter 26. Designs and Analysis of Case-Control Studies		693
26.1	Designs of Case-Control Studies	693
26.2	Analysis of Data from Design for Group Comparison	700
26.3	Analysis of Matched Data	710
26.4	Computerized Experiments	717
26.5	Practice and Experiments	719
Chapter 27. Design and Analysis of Diagnostic and Screening Tests		721
27.1	Design and Data Layout	721
27.2	Measures Frequently Used in Diagnostic Test	721
27.3	Analysis of ROC Curve	727
27.4	Decision Making on Diagnostic and Screening Test	735
27.5	Computerized Experiments	739
27.6	Practice and Experiments	742
Chapter 28. Design and Analysis of Sequential Experiments		747
28.1	Introduction	747
28.2	Design and Analysis of Sequential Trials	748
28.3	Group Sequential Schemes	755
28.4	Computerized Experiments	763
28.5	Practice and Experiments	764
Chapter 29. Systematic Review of Medical Research and Meta-Analysis		767
29.1	Basic Notions	767
29.2	Statistical Methods Commonly Used in Meta-Analysis	773
29.3	Notes	784
29.4	Computerized Experiments	787
29.5	Practice and Experiments	790
Chapter 30. Comparative Effectiveness Research		793
30.1	Background	793

30.2	Definitions	794
30.3	Examples	796
30.4	Features and Principles	799
30.5	Research Methods and Techniques	802
30.6	Steps of CER	819
30.7	Standards for Implementation and Report	822
30.8	Summary	824
30.9	Computerized Experiments	825
Chapter 31. Statistical Methods in Scale Development		831
31.1	Development of Scales	831
31.2	Adopting Scale with Foreign Language	835
31.3	The Concept and Evaluation of Validity and Reliability	840
31.4	Item Response Theory and Scale Evaluation	851
31.5	Computer Experiments	856
31.6	Exercises and Experiments	856
Chapter 32. Statistical Methods for Data from Genetic Epidemiological Study		859
32.1	Basic Concepts	859
32.2	Linkage Analysis	865
32.3	Genetic Association Analysis	871
32.4	Computerized Experiments	877
32.5	Practice and Experiments	879
Chapter 33. Statistical Methods in Bioinformatics		881
33.1	Sequence Alignment Methods	882
33.2	The Data Acquisition and Standardization of Gene Expression Patterns	885
33.3	Differentially Expressed Genes Screening	887
33.4	Cluster Analysis of Gene Expression	890
33.5	Analysis of Gene Regulatory Networks	900
33.6	Computerized Experiments	904
33.7	Summary	906
33.8	Practice and Experiment	908

Appendix I. Introduction to the Statistical Analysis System (SAS)*	
Appendix II. Statistical Tables	909
Appendix III. Datasets of Some Real Medical Examples	983
Appendix IV. Answers to Exercises*	
Appendix V. SAS Programs and Data*	

*Appendices I, IV and V are available at <http://www.worldscientific.com/r/8981-supp>, please register/sign in at the website.