

Contents

Preface xi

CHAPTER 1

Motivation i

- 1.1 Vocabulary 2
- 1.2 Population Parameters 4
- 1.3 Sampling and Sample Statistics 7
- 1.4 Statistical Inference 10

CHAPTER 2

Summarizing Data 15

- 2.1 Background 17
 - 2.1.1 Vocabulary 17
 - 2.1.2 Classification of Variables 17
 - 2.1.3 Notation 18
- 2.2 Descriptive Statistics and Graphical Methods 19
 - 2.2.1 Numerical Summaries for Continuous Variables 19
 - 2.2.2 Graphical Summaries for Continuous Variables 39
 - 2.2.3 Numerical Summaries for Discrete Variables 44
 - 2.2.4 Graphical Summaries for Discrete Variables 51
- 2.3 Key Formulas 57
 - 2.3.1 Graphical Methods: Continuous Variables 58
 - 2.3.2 Graphical Methods: Discrete Variables 59
- 2.4 Statistical Computing 59
 - 2.4.1 Continuous Variables 60
 - 2.4.2 Discrete Variables 62
 - 2.4.3 Summary of SAS Procedures 66

- 2.5 Analysis of Framingham Heart Study Data 67
- 2.6 Problems 72

CHAPTER 3**Probability 87**

- 3.1 Background 88
 - 3.1.1 Vocabulary 88
- 3.2 First Principles 90
- 3.3 Combinations and Permutations 98
- 3.4 The Binomial Distribution 103
- 3.5 The Normal Distribution 108
 - 3.5.1 Percentiles of the Normal Distribution 118
 - 3.5.2 Normal Approximation to the Binomial 120
- 3.6 Key Formulas 124
- 3.7 Applications Using SAS 125
 - 3.7.1 Summary of SAS Functions 133
- 3.8 Analysis of Framingham Heart Study Data 133
- 3.9 Problems 136

CHAPTER 4**Sampling Distributions 149**

- 4.1 Background 150
- 4.2 The Central Limit Theorem 152
- 4.3 Key Formulas 162
- 4.4 Applications Using SAS 162
 - 4.4.1 Summary of SAS Functions 167
- 4.5 Problems 167

CHAPTER 5**Statistical Inference: Procedures
for μ 173**

- 5.1 Estimating μ 174
 - 5.1.1 Vocabulary and Notation 174
 - 5.1.2 Confidence Intervals for μ 177
 - 5.1.3 Precision and Sample Size Determination 184

- 5.2 Hypothesis Tests About μ 190
 - 5.2.1 Vocabulary and Notation 190
 - 5.2.2 Hypothesis Tests About μ 196
 - 5.2.3 Power and Sample Size Determination 204
- 5.3 Key Formulas 213
- 5.4 Statistical Computing 213
 - 5.4.1 Summary of SAS Procedures and Functions 217
- 5.5 Analysis of Framingham Heart Study Data 218
- 5.6 Problems 220

CHAPTER 6

Statistical Inference: Procedures for $(\mu_1 - \mu_2)$ 231

- 6.1 Statistical Inference Concerning $(\mu_1 - \mu_2)$ 234
 - 6.1.1 Case 1: Two Independent Populations—Population Variances Known 235
 - 6.1.2 Case 2: Two Independent Populations—Population Variances Unknown but Assumed Equal 239
 - 6.1.3 Case 3: Two Independent Populations—Population Variances Possibly Unequal 244
 - 6.1.4 Case 4: Two Dependent Populations—The Data Are Matched or Paired 259
- 6.2 Power and Sample Size Determination 265
- 6.3 Key Formulas 271
- 6.4 Statistical Computing 272
 - 6.4.1 Summary of SAS Procedures and Functions 276
- 6.5 Analysis of Framingham Heart Study Data 277
- 6.6 Problems 281

CHAPTER 7

Categorical Data 293

- 7.1 Statistical Inference Concerning p 294
- 7.2 Cross-Tabulation Tables 301
- 7.3 Diagnostic Tests: Sensitivity and Specificity 303
- 7.4 Statistical Inference Concerning $(p_1 - p_2)$ 305

- 7.5 Chi-Square Tests 310
 - 7.5.1 Goodness-of-Fit Test 310
 - 7.5.2 Tests of Independence 317
- 7.6 Precision, Power, and Sample Size Determination 328
- 7.7 Key Formulas 333
- 7.8 Statistical Computing 333
 - 7.8.1 Summary of SAS Procedures 341
- 7.9 Analysis of Framingham Heart Study Data 342
- 7.10 Problems 347

CHAPTER 8

Comparing Risks in Two Populations 359

- 8.1 Effect Measures 360
- 8.2 Confidence Intervals for Effect Measures 363
- 8.3 The Chi-Square Test of Homogeneity 368
- 8.4 Fisher's Exact Test 375
- 8.5 Cochran-Mantel-Haenszel Method 377
- 8.6 Precision, Power, and Sample Size Determination 382
- 8.7 Key Formulas 383
- 8.8 Statistical Computing 384
 - 8.8.1 Summary of SAS Procedures 390
- 8.9 Analysis of Framingham Heart Study Data 391
- 8.10 Problems 400

CHAPTER 9

Analysis of Variance 407

- 9.1 Background Logic 408
- 9.2 Notation and Examples 413
- 9.3 Fixed Versus Random Effects Models 424
- 9.4 Evaluating Treatment Effects 424
- 9.5 Multiple Comparisons Procedures 425
 - 9.5.1 The Scheffe Procedure 427
 - 9.5.2 The Tukey Procedure 432
- 9.6 Repeated Measures Analysis of Variance 436
- 9.7 Key Formulas 442

- 9.8 Statistical Computing 442
 - 9.8.1 Summary of SAS Procedures 450
- 9.9 Analysis of Framingham Heart Study Data 451
- 9.10 Problems 454

CHAPTER 10**Correlation and Regression 465**

- 10.1 Correlation Analysis 466
 - 10.1.1 The Sample Correlation Coefficient r 469
 - 10.1.2 Statistical Inference Concerning p 472
- 10.2 Simple Linear Regression 477
- 10.3 Multiple Regression Analysis 485
- 10.4 Logistic Regression Analysis 487
- 10.5 Key Formulas 491
- 10.6 Statistical Computing 492
 - 10.6.1 Summary of SAS Procedures 496
- 10.7 Analysis of Framingham Heart Study Data 497
- 10.8 Problems 501

CHAPTER 11**Logistic Regression Analysis 507**

- 11.1 The Logistic Model 508
- 11.2 Statistical Inference for Simple Logistic Regression 510
- 11.3 Multiple Logistic Regression 517
- 11.4 Area under the ROC Curve 522
- 11.5 Key Formulas 527
- 11.6 Statistical Computing 527
 - 11.6.1 Summary of SAS Procedures 531
- 11.7 Analysis of Framingham Heart Study Data 631
- 11.8 Problems 538

CHAPTER 12**Nonparametric Tests 545**

- 12.1 The Sign Test (Two Dependent Samples Test) 546
- 12.2 The Wilcoxon Signed-Rank Test (Two Dependent Samples Test) 549

- 12.3 The Wilcoxon Rank Sum Test (Two Independent Samples Test) 554
- 12.4 The Kruskal-Wallis Test (k Independent Samples Test) 558
- 12.5 Spearman Correlation (Correlation Between Variables) 560
- 12.6 Key Formulas 565
 - 12.6.1 Guidelines for Determining When to Use a Nonparametric Procedure 566
- 12.7 Statistical Computing 566
 - 12.7.1 Summary of SAS Procedures 572
- 12.8 Analysis of Framingham Heart Study Data 573
- 12.9 Problems 581

CHAPTER 13

Introduction to Survival Analysis 585

- 13.1 Incomplete Follow-Up 586
- 13.2 Time to Event 588
- 13.3 Survival Analysis Techniques 589

APPENDIX A

Introduction to Statistical Computing Using SAS 591

- A.1 Introduction to SAS 591
 - A.1.1 Components of a SAS Program 591
 - A.1.2 Overview of the SAS System 593
 - A.1.3 Creating and Executing SAS Programs 595
- A.2 The Data Step 596
 - A.2.1 Inputting Data: Types of Data 597
 - A.2.2 Inputting Data: Types of Input 598
 - A.2.3 Missing Values 599
 - A.2.4 Advanced Data Input and Data Manipulation 600

APPENDIX B

Statistical Tables 609

- B.1 Statistical Tables 609
 - Table B.1 Probabilities of the Binomial Distribution 609
 - Table B.2 Probabilities of the Standard Normal Distribution 618

Table B.2A	Z Values for Confidence Intervals	620
Table B.2B	Z Values for Tests of Hypothesis	620
Table B.3	Critical Values of the t Distribution	621
Table B.4A	Critical Values of the F Distribution with Upper-Tail Area = 0.05	624
Table B.4B	Critical Values of the F Distribution with Upper-Tail Area = 0.205	625
Table B.5	Critical Values of the χ^2 Distribution	626
Table B.6	Critical Values of the Studentized Range Distribution, $\alpha = 0.05$	627
B.2	SAS Programs Used to Generate Table Entries	628

APPENDIX C

Framingham Heart Study Longitudinal Data Documentation 631

Index 645

CHAPTER 11

Logistic Regression Analysis 507

11.1	The Logistic Model	507
11.2	Statistical Inference for Logistic Regression	511
11.3	Multiple Logistic Regression	517
11.4	Area under the ROC Curve	525
11.5	Key Formulas	527
11.6	Statistical Computing	527
11.6.1	Summary of SAS Programs	527
11.7	Analysis of Framingham Heart Study Data	529
11.8	Problems	535

CHAPTER 12

Nonparametric Tests 509

12.1	The Sign Test (Two Dependent Samples)	509
12.2	The Wilcoxon Signed-Rank Test	511
12.3	The Wilcoxon Rank-Sum Test	517
12.4	Key Formulas	527
12.5	Statistical Computing	527
12.6	Analysis of Framingham Heart Study Data	529
12.7	Problems	535