

STATISTICAL CONCEPTS AND METHODS



Gouri K. Bhattacharyya
Richard A. Johnson

Contents

1	INTRODUCTION	1
1.1	What Is Statistics?	1
1.2	Statistics and Everyday Life,	2
1.3	Elements of Scientific Inquiry,	3
1.4	The Role of Statistics in Scientific Inquiry,	4
1.5	Illustrative Situations of the Collection and Analysis of Data,	5
1.6	Population and Sample,	6
1.7	Statistics Interacting with Other Fields,	9
	References,	10
2	DESCRIPTIVE STUDY OF DATA	11
2.1	Introduction,	11
2.2	Description of Data by Graphs and Tables,	12
2.3	Symbols for Data Sets and the Operation of Addition, 24	
2.4	Measures of Center,	27
2.5	Measures of Variation,	32
2.6	Comments on the Choice of Numerical Measures,	38
2.7	Descriptive Measures for Grouped Data,	40
*2.8	Coding the Data,	45
	Exercises,	47
	Class Projects,	57
	Mathematical Exercises,	58
3	ELEMENTS OF PROBABILITY	60
3.1	Introduction,	60
3.2	Sample Space and Events,	61

x

- 3.3 The Probability of an Event, 64
- 3.4 Operations with Events and Basic Laws of Probability, 70
- 3.5 Rules of Counting and Their Use in Uniform Probability Models, 78
- 3.6 Conditional Probability, 87
- *3.7 Bayes' Theorem, 93
- *3.8 General Rules of Probability, 95
- Exercises, 97
- Class Projects, 106
- Mathematical Exercises, 107

4 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

108

- 4.1 Random Variables, 108
- 4.2 Graphic Presentation of a Probability Distribution, 113
- 4.3 Expectation and Its Properties, 117
- 4.4 Variance: A Measure of Spread, 121
- 4.5 Joint Distribution of Two Random Variables, 124
- 4.6 Covariance and Correlation, 129
- 4.7 Independence of Two Random Variables, 131
- Exercises, 133
- Mathematical Exercises, 138

5 DISTRIBUTIONS FOR COUNTS

140

- 5.1 Idea of a Probability Model, 140
- 5.2 Bernoulli Trials: Success-Failure, 141
- 5.3 The Binomial Distribution, 145
- 5.4 How to Use the Binomial Tables, 150
- 5.5 The Mean and the Variance of the Binomial Distribution, 151
- 5.6 The Hypergeometric Distribution, 152
- 5.7 The Geometric Distribution for Waiting Time, 154
- 5.8 Rare Events and the Poisson Distribution, 156
- References, 159
- Exercises, 159
- Mathematical Exercises, 163

6 BASIC CONCEPTS OF TESTING HYPOTHESES

165

- 6.1 Introduction, 165
- 6.2 The Null and the Alternative Hypotheses, 166

6.3	The Two Types of Errors and the Power Function of a Test, 169	
6.4	Selecting Among Several Tests, 172	
6.5	Drawing Conclusions from a Test, 174	
6.6	Testing with a Two-sided Alternative, 176	
6.7	General Steps in Testing Hypotheses, 178	
	Exercises, 182	
7	THE NORMAL DISTRIBUTION AND RANDOM SAMPLES	187
7.1	Probability Models for Continuous Variables, 187	
7.2	The Normal Distribution, 193	
7.3	The Normal Approximation to the Binomial Distribution, 202	
7.4	Random Sampling, Statistic, and Sampling Distribution, 205	
7.5	Distribution of the Sample Mean and the Central Limit Theorem, 210	
7.6	Checking the Assumption of a Normal Population, 220	
7.7	Transforming Observations to Attain Near Normality, 223	
	References, 226	
	Exercises, 226	
	Class Projects, 232	
8	INFERENCES ABOUT A POPULATION	233
8.1	Introduction, 233	
8.2	Point Estimation of a Parameter, 235	
8.3	Estimation by Confidence Intervals, 243	
8.4	Confidence Interval for μ Based on Small Samples (t Distribution), 250	
8.5	Testing Hypotheses About a Mean, 255	
8.6	Relationship Between Tests and Confidence Intervals, 264	
8.7	Inferences About the Standard Deviation σ of a Normal Population, 266	
8.8	The Effect of Violation of Assumptions, 270	
8.9	Determination of the Sample Size, 272	
	Reference, 275	
	Exercises, 276	
	Mathematical Exercises, 283	

9 COMPARING TWO TREATMENTS

- 9.1 Introduction, 286
- 9.2 Independent Random Samples from Two Populations, 288
- 9.3 Randomization and Its Role in Inference, 297
- 9.4 Paired Comparisons, 300
- 9.5 Choosing Between Independent Samples and Paired Samples, 306
- 9.6 Comparison of Two Binomial Proportions, 308
- 9.7 Comparing the Variances of Two Normal Populations, 312
- 9.8 Relationships of χ^2 , t , and F to the Normal Distribution, 317
- Exercises, 321

10 REGRESSION ANALYSIS: SIMPLE LINEAR REGRESSION

- 10.1 Objectives of Regression Analysis, 334
- 10.2 A Simple Regression Problem, 338
- 10.3 Plotting the Data, 339
- 10.4 A Straight-line Regression Model, 341
- 10.5 The Method of Least Squares, 342
- 10.6 Derivation of the Least Squares Estimators, 346
- 10.7 Least Squares Regression Line, 348
- 10.8 Properties of the Least Squares Estimators, 349
- 10.9 Important Inferential Problems, 351
- 10.10 What to Do if the Predictor Variable Cannot be Controlled by the Experimenters, 357
- 10.11 Remarks About the Model, 358
- Exercises, 359
- Mathematical Exercises, 366

11 REGRESSION ANALYSIS: MODEL CHECKING AND MULTIPLE LINEAR REGRESSION

- 11.1 General Attitude Toward a Statistical Model, 368
- 11.2 Examining the Residuals, 368
- 11.3 Further Checks on the Straight-line Model, 372
- 11.4 Test of Lack of Fit Based on Repeated Runs, 373
- 11.5 Nonlinear Relations and Linearizing Transformations, 379
- 11.6 Multiple Linear Regression, 384

*11.7	Further Aspects of Multiple Regression, 388	
	References, 391	
	Exercises, 391	
	Mathematical Exercises, 399	
12	CORRELATION: A MEASURE OF LINEAR RELATIONSHIP	400
12.1	Introduction, 400	
12.2	The Sample Correlation Coefficient, 402	
12.3	Correlation and Causation, 406	
12.4	A Population Model for Correlation: The Bivariate Normal, 408	
12.5	Serial Correlation, 411	
	Exercises, 413	
	Mathematical Exercises, 418	
13	ANALYSIS OF CATEGORIZED DATA	420
13.1	Introduction, 420	
13.2	The Multinomial Model, 422	
13.3	Pearson's Test for Goodness of Fit, 424	
13.4	Contingency Tables, 428	
13.5	Contingency Tables with One Margin Fixed (Test of Homogeneity), 435	
13.6	An Exact Test for a 2×2 Contingency Table, 441	
	References, 444	
	Exercises, 444	
	Mathematical Exercises, 451	
14	DESIGN OF EXPERIMENTS AND ANALYSIS OF VARIANCE	453
14.1	Introduction and Basic Terminology, 453	
14.2	Comparison of Several Treatments: The Completely Randomized Design, 456	
14.3	Population Model and Inferences for a Completely Randomized Design, 462	
14.4	Randomized Block Experiments for Comparing k Treatments, 469	
14.5	Factorial Experiments, 477	
14.6	Two-level Factorial Designs, 483	
	References, 496	
	Exercises, 496	
	Mathematical Exercises, 503	

15 NONPARAMETRIC INFERENCE

- 15.1 Introduction, 505
- 15.2 The Wilcoxon Rank-Sum Test for Comparing Two Treatments, 507
- 15.3 Paired Comparisons, 516
- 15.4 Confidence Intervals Based on Rank Tests, 523
- 15.5 Measures of Correlation Based on Ranks, 526
- 15.6 The Kruskal-Wallis Test for Comparing k Treatments, 533
- 15.7 Concluding Remarks, 537
- References, 539
- Exercises, 539
- Class Project, 547
- Mathematical Exercises, 547

16 SAMPLE SURVEYS

- 16.1 Why Take Only a Sample When the Population Is Finite? 550
- 16.2 Specification of the Population and the Characteristic of Interest, 551
- 16.3 Probability Sampling, 552
- 16.4 Bias and Its Sources, 552
- 16.5 Using a Random Number Table, 554
- 16.6 Simple Random Sampling, 555
- 16.7 Sampling to Determine a Proportion, 560
- 16.8 Stratified Random Sampling, 563
- 16.9 Allocation of Sample Sizes, 569
- 16.10 Stratified Sampling to Determine a Proportion, 571
- 16.11 Other Methods of Sampling, 573
- 16.12 Planning a Survey, 575
- References, 577
- Exercises, 577
- Class Projects, 583
- Mathematical Exercises, 583

APPENDIX

- Table 1 The Number of Combinations $\binom{n}{r}$, 587
- Table 2 Cumulative Binomial Probabilities, 588
- Table 3 Cumulative Poisson Probabilities, 595

Table 4	Standard Normal Probabilities, 597
Table 5	Percentage Points of t Distributions, 599
Table 6	Percentage Points of χ^2 Distributions, 600
Table 7	Percentage Points of $F(p_1, p_2)$ Distributions, 602
Table 8	Extended Table of Percentage Points of t Distributions, 608
Table 9	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Rank-Sum Statistic, 610
Table 10	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Signed-Rank Statistic, 615
Table 11	Confidence Intervals for the Median of a Symmetric Distribution Using the Wilcoxon Signed-Rank Test, 617
Table 12	Confidence Intervals for Shift Using the Wilcoxon Rank-Sum Test, 618
Table 13	Selected Tail Probabilities for the Null Distribution of Spearman's Statistic $\sum_{i=1}^n R_i S_i$, 621
Table 14	5000 Random Digits, 623
	Answers to Selected Odd Numbered Exercises, 625

INDEX

631