

Contents

Preface	xv
1. What is Statistics? How is it Applied in the Health Sciences?	1
1.1 Definitions of Statistics and Statisticians	2
1.2 Why Study Statistics?	3
1.3 Types of Studies	8
1.3.1 Surveys and Cross-Sectional Studies	9
1.3.2 Retrospective Studies	10
1.3.3 Prospective Studies	10
1.3.4 Experimental Studies and Quality Control	10
1.3.5 Clinical Trials	12
1.3.6 Epidemiological Studies	14
1.3.7 Pharmacoeconomic Studies and Quality of Life	16
1.4 Exercises	18
1.5 Additional Reading	19
2. Defining Populations and Selecting Samples	22
2.1 What are Populations and Samples?	22
2.2 Why Select a Sample?	23
2.3 How Samples Can be Selected	25
2.3.1 Simple Random Sampling	25
2.3.2 Convenience Sampling	25
2.3.3 Systematic Sampling	26
2.3.4 Stratified Random Sampling	28
2.3.5 Cluster Sampling	28
2.3.6 Bootstrap Sampling	29

- 2.4 How to Select a Simple Random Sample 29
- 2.5 How to Select a Bootstrap Sample 39
- 2.6 Why Does Random Sampling Work? 41
- 2.7 Exercises 41
- 2.8 Additional Reading 45

- 3. Systematic Organization and Display of Data 46**
 - 3.1 Types of Data 46
 - 3.1.1 Qualitative 47
 - 3.1.2 Quantitative 47
 - 3.2 Frequency Tables and Histograms 48
 - 3.3 Graphical Methods 51
 - 3.3.1 Frequency Histograms 51
 - 3.3.2 Frequency Polygons 53
 - 3.3.3 Cumulative Frequency Polygon 54
 - 3.3.4 Stem-and-Leaf Diagrams 56
 - 3.3.5 Box-and-Whisker Plots 58
 - 3.3.6 Bar Charts and Pie Charts 61
 - 3.1 Exercises 63
 - 3.1 Additional Reading 67

- 4. Summary Statistics 68**
 - 4.1 Measures of Central Tendency 61
 - 4.1.1 The Arithmetic Mean 68
 - 4.1.2 The Median 70
 - 4.1.3 The Mode 73
 - 4.1.4 The Geometric Mean 73
 - 4.1.5 The Harmonic Mean 74
 - 4.1.6 Which Measure Should You Use? 75
 - 4.2 Measures of Dispersion 76
 - 4.2.1 Range 78
 - 4.2.2 Mean Absolute Deviation 78
 - 4.2.3 Population Variance and Standard Deviation 79
 - 4.2.4 Sample Variance and Standard Deviation 82
 - 4.2.5 Calculating the Variance and Standard Deviation from Group Data 84
 - 4.3 Coefficient of Variation (CV) and Coefficient of Dispersion (CD) 85
 - 4.4 Exercises 88
 - 4.5 Additional Reading 91

5. Basic Probability	92
5.1 What is Probability?	92
5.2 Elementary Sets as Events and Their Complements	95
5.3 Independent and Disjoint Events	95
5.4 Probability Rules	98
5.5 Permutations and Combinations	100
5.6 Probability Distributions	103
5.7 The Binomial Distribution	109
5.8 The Monty Hall Problem	110
5.9 A Quality Assurance Problem	113
5.10 Exercises	115
5.11 Additional Reading	120
6. The Normal Distribution	121
6.1 The Importance of the Normal Distribution in Statistics	121
6.2 Properties of Normal Distributions	122
6.3 Tabulating Areas under the Standard Normal Distribution	124
6.4 Exercises	129
6.5 Additional Reading	132
7. Sampling Distributions for Means	133
7.1 Population Distributions and the Distribution of Sample Averages from the Population	133
7.2 The Central Limit Theorem	141
7.3 Standard Error of the Mean	143
7.4 Z Distribution Obtained When Standard Deviation Is Known	144
7.5 Student's t Distribution Obtained When Standard Deviation Is Unknown	144
7.6 Assumptions Required for t Distribution	147
7.7 Exercises	147
7.8 Additional Reading	149
8. Estimating Population Means	150
8.1 Estimation Versus Hypothesis Testing	150
8.2 Point Estimates	151
8.3 Confidence Intervals	153
8.4 Confidence Intervals for a Single Population Mean	154
8.5 Z and t Statistics for Two Independent Samples	159

8.6	Confidence Intervals for the Difference between Means from Two Independent Samples (Variance Known)	161
8.7	Confidence Intervals for the Difference between Means from Two Independent Samples (Variance Unknown)	161
8.8	Bootstrap Principle	166
8.9	Bootstrap Percentile Method Confidence Intervals	167
8.10	Sample Size Determination for Confidence Intervals	176
8.11	Exercises	179
8.12	Additional Reading	181
9.	Tests of Hypotheses	182
9.1	Terminology	182
9.2	Neyman–Pearson Test Formulation	183
9.3	Test of a Mean (Single Sample, Population Variance Known)	186
9.4	Test of a Mean (Single sample, Population Variance Unknown)	187
9.5	One-Tailed Versus Two-Tailed Tests	188
9.6	p -Values	191
9.7	Type I and Type II Errors	191
9.8	The Power Function	192
9.9	Two-Sample t Test (Independent Samples with a Common Variance)	193
9.10	Paired t Test	195
9.11	Relationship between Confidence Intervals and Hypothesis Tests	199
9.12	Bootstrap Percentile Method Test	200
9.13	Sample Size Determination for Hypothesis Tests	201
9.14	Sensitivity and Specificity in Medical Diagnosis	202
9.15	Meta-Analysis	204
9.16	Bayesian Methods	207
9.17	Group Sequential Methods	209
9.18	Missing Data and Imputation	210
9.19	Exercises	212
9.20	Additional Reading	215
10.	Inferences Regarding Proportions	217
10.1	Why Are Proportions Important?	217
10.2	Mean and Standard Deviation for the Binomial Distribution	218
10.3	Normal Approximation to the Binomial	221
10.4	Hypothesis Test for a Single Binomial Proportion	222
10.5	Testing the Difference between Two Proportions	224

10.6	Confidence Intervals for Proportions	225
10.7	Sample Size Determination—Confidence Intervals and Hypothesis Tests	227
10.8	Exercises	228
10.9	Additional Reading	229
11.	Categorical Data and Chi-Square Tests	231
11.1	Understanding Chi-Square	232
11.2	Chi-Square Distributions and Tables	233
11.3	Testing Independence between Two Variables	233
11.4	Testing for Homogeneity	236
11.5	Testing for Differences between Two Proportions	237
11.6	The Special Case of 2×2 Contingency Table	238
11.7	Simpson's Paradox in the 2×2 Table	239
11.8	McNemar's Test for Correlated Proportions	241
11.9	Relative Risk and Odds Ratios	242
11.10	Goodness of Fit Tests—Fitting Hypothesized Probability Distributions	244
11.11	Limitations to Chi-Square and Exact Alternatives	246
11.12	Exercises	247
11.13	Additional Reading	250
12.	Correlation, Linear Regression, and Logistic Regression	251
12.1	Relationships between Two Variables	252
12.2	Uses of Correlation and Regression	252
12.3	The Scatter Diagram	254
12.4	Pearson's Product Moment Correlation Coefficient and Its Sample Estimate	256
12.5	Testing Hypotheses about the Correlation Coefficient	258
12.6	The Correlation Matrix	259
12.7	Regression Analysis and Least Squares Inference Regarding the Slope and Intercept of a Regression Line	259
12.8	Sensitivity to Outliers, Outlier Rejection, and Robust Regression	264
12.9	Galton and Regression toward the Mean	271
12.10	Multiple Regression	277
12.11	Logistic Regression	283
12.12	Exercises	287
12.13	Additional Reading	293

13. One-Way Analysis of Variance	295
13.1 Purpose of One-Way Analysis of Variance	296
13.2 Decomposing the Variance and Its Meaning	297
13.3 Necessary Assumptions	298
13.4 <i>F</i> Distribution and Applications	298
13.5 Multiple Comparisons	301
13.5.1 General Discussion	301
13.5.2 Tukey's Honest Significant Difference (HSD) Test	301
13.6 Exercises	302
13.7 Additional Reading	307
14. Nonparametric Methods	308
14.1 Advantages and Disadvantages of Nonparametric Versus Parametric Methods	308
14.2 Procedures for Ranking Data	309
14.3 Wilcoxon Rank-Sum Test	311
14.4 Wilcoxon Signed-Rank Test	314
14.5 Sign Test	317
14.6 Kruskal–Wallis Test: One-Way ANOVA by Ranks	319
14.7 Spearman's Rank-Order Correlation Coefficient	322
14.8 Permutation Tests	324
14.8.1 Introducing Permutation Methods	324
14.8.2 Fisher's Exact Test	327
14.9 Insensitivity of Rank Tests to Outliers	330
14.10 Exercises	331
14.11 Additional Reading	334
15. Analysis of Survival Times	336
15.1 Introduction to Survival Times	336
15.2 Survival Probabilities	338
15.2.1 Introduction	338
15.2.2 Life Tables	339
15.2.3 The Kaplan–Meier Curve	341
15.2.4 Parametric Survival Curves	344
15.2.5 Cure Rate Models	348
15.3 Comparing Two or More Survival Curves—The Log Rank Test	349
15.4 Exercises	352
15.5 Additional Reading	354

16. Software Packages for Statistical Analysis	356
16.1 General-Purpose Packages	356
16.2 Exact Methods	359
16.3 Sample Size Determination	359
16.4 Why You Should Avoid Excel	360
16.5 References	361
Postscript	362
Appendices	363
A Percentage Points, <i>F</i> -Distribution ($\alpha = 0.05$)	363
B Studentized Range Statistics	364
C Quantiles of the Wilcoxon Signed-Rank Test Statistic	366
D χ^2 Distribution	368
E Table of the Standard Normal Distribution	370
F Percentage Points, Student's <i>t</i> Distribution	371
G Answers to Selected Exercises	373
Index	401