
Contents

<i>Preface</i>	xv
Basics	
1 Statistical Distributions	3
The Normal Probability Distribution	3
The t -Distribution	7
The Chi-Square Probability Distribution	9
Illustration of the Pearson Chi-Square Statistic – Simplest Case	12
The f -Distribution	13
The Uniform Probability Distribution	13
The p -value	15
A Few Relationships among Probability Distributions	16
2 Confidence Intervals	19
Four Properties of an Estimated Confidence Interval	20
Confidence Intervals for a Function of an Estimate	21
Confidence Intervals Based on Estimates near Zero	21
Exact Confidence Interval (Computer-Estimated)	22
A Confidence Interval Based on an Estimated Median Value	23
A Confidence Interval and a Confidence Band	25
Details: A Confidence Interval and a Statistical Test	27
3 A Weighted Average	29
A Basic Application	29
Ratios and Weighted Averages	30
Estimates Weighted by Reciprocal Variances	32
A Puzzle	37
Age-Adjusted Rates Using Weighted Averages	39
Smoothing – A Weighted Average Approach	40
Example: Weighted Average Smoothing of Hodgkin’s Disease Mortality Data	44
4 Two Discrete Probability Distributions	47
Binomial Probability Distribution	49

	Two Applications of the Binomial Distribution	53
	The Geometric Probability Distribution	57
	A Poisson Probability Distribution	58
	Two Applications of the Poisson Probability Distribution	62
	A Note on Rare Events	67
5	Correlation	69
	Spearman's Rank Correlation Coefficient	73
	Point Biserial Correlation Coefficient	74
	Nonparametric Measure of Association: The γ -Coefficient	77
	A Special Case: The $2 \times k$ Table	80
	Chi-Square-Based Measures of Association	81
	Proportional Reduction in Error Criterion	83
	Applications	
6	The 2×2 Table	87
	The Analysis of a 2×2 Table	89
	Measures of Association in a 2×2 Table	92
	Odds Ratio and Relative Risk Ratio	94
	A Correction to Improve the Normal Distribution Approximation	96
	The Hypergeometric Probability Distribution	98
	Fisher's Exact Test	101
	Correlation in a 2×2 Table	102
	A 2×2 Table with a Structural Zero	106
	Assessing the Accuracy of a Diagnostic Test	107
7	Linear Bivariate Regression Model	111
	The Bivariate Linear Regression Model	111
	Additivity	114
	Coefficients	114
	Multiple Correlation Coefficient	116
	Adjustment	118
	Interaction	119
	Confounder Bias	121
	Collinearity	123
8	The $2 \times k$ Table	126
	Wilcoxon (Mann-Whitney) Rank Sum Test	126
	Nonparametric Analysis of a $2 \times k$ Table	132
	A Chi-Square Analysis of a $2 \times k$ Table	134
	Another Example: Childhood Cancer and X-Ray Exposure	137
9	The Loglinear Poisson Regression Model	141
	A Simple Poisson Regression Model	141

Contents

ix

	Poisson Regression Model: Analysis of Vital Statistics Data	144
	Rate Ratios Adjusted for Several Variables	146
	A Test for Linear Trend	148
	A Graphical Display of a Table	150
	Implementation of Poisson Models Applied to Tabular Data	151
	Analysis of Tables with Incomplete Data	153
	Another Illustration of the Analysis of an Incomplete Table	154
	Quasi-independence: Analysis of an Incomplete Table	156
	A Case Study – Adjusted Mortality Rates: Black-White Infant Mortality	158
	First Approach: Weight-Specific Comparisons	161
	Second Approach: A Model-Free Summary	163
	Third Approach: Poisson Regression Model	165
10	Two-Way and Three-Way Tables and Their Analysis	170
	Analysis of Tables – Continuous Data	177
	Matched Pairs – The Categorical Variable Case	182
	Analysis of Tables – Count Data	185
	Three-Way Tables	189
	The Analysis of the Three-Way Table	190
	Complete Independence	190
	Joint Independence	191
	Conditional Independence	191
	No Pairwise Independence	192
	Log-Linear Models for Four $2 \times 2 \times 2$ Three-Way Tables	193
	Example of Completely Independent Variables	193
	Example of Jointly Independent Variables	194
	Example of Conditional Independent Variables	195
	Example of Additive Relationships	196
	A Case Study – Joint Independence	197
	A Case Study – Conditional Independence in a $2 \times 2 \times 4$ Table	201
11	Bootstrap Analysis	204
	Example: Analysis of Paired Data	209
	Example: Evaluation of a Difference between Two Harmonic Mean Values	211
	An Example of Bootstrap Estimation from Categorical Data	214
	Kappa Statistic – A Bootstrap-Estimated Confidence Interval	216
	A Graphic Application of Bootstrap Estimation	218
	A Property of Bootstrap Estimation	220
	Randomization and Bootstrap Analysis Applied to Two-Sample Data	220
	Randomization Test	222
	Bootstrap Analysis	223

12	Graphical Analysis	227
	A Confidence Interval and a Boxplot	227
	Multiple Comparisons – A Visual Approach	229
	The Cumulative Probability Distribution Function	229
	Inverse Functions for Statistical Analysis	232
	Kolmogorov One-Sample Test	241
	Kolmogorov-Smirnov Two-Sample Test	243
	Simulation of Random “Data” with a Specific Probability Distribution	245
	Simulation of “Data” with a Continuous Probability Distribution	248
	Simulation of “Data” with a Discrete Probability Distribution	249
	A Case Study – Poisson Approximation	249
	A Graphical Approach to Regression Analysis Diagnostics	250
	A Graphical Goodness-of-Fit for the Logistic Regression Model	257
13	The Variance	260
	A Brief Note on Estimation of the Variance	260
	A Confidence Interval	261
	Test of Variance	262
	Homogeneity/Heterogeneity	264
	Analysis of Variance – Mean Values	265
	Another Partitioning of Variability	268
	Two-Sample Test of Variance	269
	<i>F</i> -Ratio Test of Variance	270
	Bartlett’s Test of Variance	271
	Levene’s Test of Variance	273
	Siegel-Tukey Two-Sample Test of Variance	274
	Comparison of Several Variances	275
14	The Log-Normal Distribution	278
	Example: Lognormal Distributed Data	281
	A Left-Censored Log-Normal Distribution	283
	An Applied Example	284
15	Nonparametric Analysis	287
	The Sign Test	289
	The Wilcoxon Signed Rank Test	291
	Kruskal-Wallis Nonparametric Comparison of <i>k</i> Sample Mean Values	294
	Three-Group Regression Analysis	297
	Tukey’s Quick Test	300
	Friedman Rank Test	302
	Survival	
16	Rates	309
	An Average Mortality Rate	309

<i>Contents</i>		xi
	An Approximate Average Rate	314
	A Rate Estimated from Continuous Survival Times	318
17	Nonparametric Survival Analysis	324
	Cumulative Hazard Function	333
	Description of the Median Survival Time	334
	Comparison of Two Survival Probability Distributions – The Log-Rank Test	335
	Proportional Hazards Rates	339
	An Example of Survival Analysis	342
	The Cox Analysis of Proportional Hazards Models	345
	Proportional Hazards Model – A Case Study	346
18	The Weibull Survival Function	352
	Two-Sample Comparison – The Weibull Model	358
	The Shape Parameter	360
	Multivariable Weibull Survival Time Model – A Case Study	363
Epidemiology		
19	Prediction, a Natural Measure of Performance	371
	Net Reclassification Index (Binary Case)	373
	Net Reclassification Index (Extended Case)	375
	Integrated Discrimination Improvement	379
	Summary Lines as Measures of Improvement in Model Prediction	380
	Covariance and Correlation	382
	Epilogue	383
	Least Squares Estimation – Details	383
20	The Attributable Risk Summary	387
	Incidence Rates	391
	Two Risk Factors	393
	Adjustment by Stratification	394
	Adjustment by Model	395
	Issues of Interpretation	396
	A Few Less Specific Issues of Interpretation	397
21	Time-Space Analysis	399
	Knox’s Time-Space Method	399
	The Statistical Question Becomes: Is $m = 0$?	400
	Test of Variance Applied to Spatial Data	403
	Mantel’s Time-Space Regression Method	404
	Performance and Time-Space Methods	405

22	ROC Curve and Analysis	407
	An ROC Curve	407
	An ROC Analysis Example	414
	Nonparametric Estimation of an ROC Curve	417
	An Example – Nonparametric ROC Analysis	418
	How to Draw a Nonparametric ROC Curve	419
	Construction of the ROC Curve	419
	Implementation	425
	Genetics	
23	Selection: A Statistical Description	433
	Stable Equilibrium	434
	A Description of Recombination	435
	Two Examples of Statistical Analyses	439
	Selection – Recessive Lethal	442
	A More General Selection Pattern	444
	Selection – A Balanced Polymorphism	445
	Fitness	448
24	Mendelian Segregation Analysis	450
	Ascertainment	450
	Truncation	451
	Estimation of a Segregation Probability: Complete Ascertainment	451
	Estimation of a Segregation Probability: Single Ascertainment	455
25	Admixed Populations	458
	A Model Describing Admixture	459
	Estimation of the Admixture Rate	461
	An Example: Estimation of the Extent of Admixture	462
26	Nonrandom Mating	465
	Genotype Frequencies – Correlation and Variance	465
	Genetic Variance	469
	Wahlund's Model	470
	Random Genetic Drift	473
	A Description: Consecutive Random Sampling of a Genetic Population	473
	Random Genetic Drift – A Description	474
	The Statistics of Random Genetic Drift	474
	Mutation/Selection Equilibrium	477
	The Story of Mr. A and Mr. B	477
	One-Way Mutation	479
	Mutation/Selection Balance	480

	<i>Contents</i>	xiii
	Assortative Mating	481
	Assortative Mating Model	482
	Theory	
27	Statistical Estimation	489
	The Sample	489
	Covariance	490
	The Population	492
	Infinite Series with Statistical Applications	496
	Example of Infinite Power Series	497
	Binomial Theorem	498
	Functions of Estimates	499
	Approximate Values for the Expectation and Variance of a Function	500
	Partitioning the Total Sum of Squares	503
	Expected Value and Variance of Wilcoxon Signed Rank Test	504
	Maximum Likelihood Estimation	505
	Properties of a Maximum Likelihood Estimate	510
	Example Maximum Likelihood Estimates	512
	Method of Moments Estimation	515
	<i>Appendix: R Code</i>	519
	<i>Index</i>	557