

# BIostatISTICS FOR MEDICAL AND BIOMEDICAL PRACTITIONERS

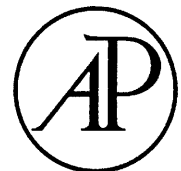
**JULIEN I. E. HOFFMAN**

Tiburon, California, USA



Amsterdam • Boston • Heidelberg • London • New York • Oxford  
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

Academic Press is an imprint of Elsevier



# CONTENTS

<i>About the Author</i>	<i>xvii</i>
<i>Preface</i>	<i>xix</i>
<i>Acknowledgments</i>	<i>xxv</i>

## **Part 1: Basic Aspects of Statistics** **1**

### **1. Basic Concepts** **3**

Introduction	3
Populations and Samples	4
The Effects of Variability	4
Variables and Parameters	5
Basic Uses of Statistics	6
Description	6
Statistical Inference	7
Data	9
Models	10
General Approach to Study Design	10
A Brief History of Statistics	13
References	16

### **2. Statistical Use and Misuse in Scientific Publications** **17**

Early Use of Statistics	17
Current Tests in Common Use	18
Statistical Misuse	19
Basic Guides to Statistics	22
References	23

### **3. Some Practical Aspects** **27**

Statistics Programs	27
Variables	29
Measurement Scales	29
Displaying Data Sets	33
Practical Issues of Data Sets	34
Accuracy of Measurement	35
Rounding off and Truncation	36
Accuracy and Precision	37

Notation	37
Operators	38
Weights	40
Statistics Books	41
References	42
<b>4. Exploratory Descriptive Analysis</b>	<b>43</b>
Basic Concepts	44
Counting	44
Distribution	45
Shapes of Distributions	50
Stem-and-Leaf Diagrams	51
Measures of Central Tendency (Location)	57
Measures of Variability	64
Tables and Graphs	70
Box Plots	74
Advanced and Alternative Concepts	76
Classical versus Robust Methods	76
Propagation of Errors	79
Three Useful Applications of the Variance	81
Appendix	82
Least Squares Principle	82
References	82
<b>5. Basic Probability</b>	<b>85</b>
Introduction	85
Types of Probability	85
Basic Principles and Definitions	86
Additional Definitions	88
Conditional Probability	91
Bayes' Theorem	93
References	98
<b>Part 2: Continuous Distributions</b>	<b>99</b>
<b>6. Normal Distribution</b>	<b>101</b>
Introduction	101
Normal or Gaussian Curve	102
The Quincunx	104
Properties of the Normal Curve	104
Populations and Samples	108

Description of the Distribution Shape	109
Skewness	110
Kurtosis	111
Determining Normality	111
Ungrouped Data	115
How Important Is Normality?	118
References	119
<b>7. Statistical Limits and the Central Limit Theorem</b>	<b>121</b>
Central Limit Theorem	121
Generalizations Based on a Single Sample	124
Tolerance Limits	127
Reporting Results	129
Graphic Representation	129
References	131
<b>8. Other Continuous Distributions</b>	<b>133</b>
Continuous Uniform Distribution	133
Exponential Distribution	133
Logarithmic Distribution	135
Weibull Distribution	136
Chi-Square Distribution	137
Variance Ratio (F) Distribution	139
References	140
<b>9. Outliers and Extreme Values</b>	<b>141</b>
Outliers	141
Grubb's Test	142
Slippage and Contamination	143
Robust Tests	144
Dealing with Outliers	144
Extreme Values	145
Appendix	147
References	148
<b>Part 3: Hypothesis Testing</b>	<b>151</b>
<b>10. Hypothesis Testing: The Null Hypothesis, Significance, and Type I Error</b>	<b>153</b>
Hypotheses	153
Significance	153
References	158

<b>11. Hypothesis Testing: Sample Size, Effect Size, Power, and Type II Errors</b>	<b>159</b>
Basic Concepts	159
Statistical Power	159
Effect Size	161
Calculation of Power	162
Post hoc Power Analysis	168
Advanced and Alternative Concepts	169
Hypothesis Testing and Maximum Likelihood; The Bayes Factor	169
References	170
 <b>Part 4: Discrete and Categorical Distributions</b>	 <b>173</b>
<b>12. Permutations and Combinations</b>	<b>175</b>
Permutations	175
Combinations	176
References	178
 <b>13. Hypergeometric Distribution</b>	 <b>179</b>
Introduction	179
General Formula	179
Fisher's Exact Test	180
Multiple Groups	182
Reference	182
 <b>14. Categorical and Cross-Classified Data: Goodness of Fit and Association</b>	 <b>183</b>
Basic Concepts	183
Introduction	183
Goodness of Fit	184
Degrees of Freedom	187
2 × 2 Contingency Tables	187
Odds Ratio	193
Cautionary Tales	195
Larger Contingency Tables	201
Fisher's Exact Test	203
Determining Power and Sample Size	206
Advanced Concepts	208
Cochran—Mantel—Haenszel Test and Confounders	208
Pooling Data	211
Problems with $R \times C$ Tables	213
Multidimensional Tables	214
References	217

<b>15. Categorical and Cross-Classified Data: McNemar's Test, Kolmogorov–Smirnov Tests, Concordance</b>	<b>221</b>
Paired Samples: McNemar's Test	221
Bowker's Test	224
Testing Ordered Categorical Data: Kolmogorov–Smirnov Tests	225
K–S One-Sample Test	226
K–S Two-Sample Test	228
Concordance (Agreement) between Observers	229
Two Categories	229
More than Two Categories	231
Weighted Kappa	232
Cautionary Tales	235
Intraclass Correlation	236
References	237
 <b>16. Binomial and Multinomial Distributions</b>	 <b>239</b>
Basic Concepts	239
Introduction	239
Binomial Basics	244
The Normal Approximation	245
Fitting a Binomial Distribution to a Set of Trial Results	245
Cumulative Binomial Probabilities	245
Confidence Limits	247
Comparing Two Binomial Distributions	249
Estimating Sample Size	249
Comparing Probabilities	250
Advanced or Alternative Concepts	250
Exact Confidence Limits	250
Multinomial Distribution	251
Appendix	252
References	252
 <b>17. Proportions</b>	 <b>253</b>
Introduction	253
Proportions and Binomial Theorem	253
Confidence Limits	254
Sample and Population Proportions	255
Sample Size	256
Comparing Proportions	257
Pooling Samples	258
References	258

<b>18. The Poisson Distribution</b>	<b>259</b>
Introduction	259
Relationship to the Binomial Distribution	261
Goodness of Fit to a Poisson Distribution	262
The Ratio of the Variance to the Mean of a Poisson Distribution	265
Setting Confidence Limits	266
Normal Approximation	266
Exact Method	267
The Square Root Transformation	268
Cumulative Poisson Probabilities	269
Differences between Means of Poisson Distributions	271
Comparison of Counts Based on Same Units	271
Comparison of Counts Not Based on Same Units	272
Comparing the Ratio of Two Poisson Variates	274
Determining the Required Sample Size	275
Waiting Times	277
Appendix	277
References	278
<b>19. Negative Binomial Distribution</b>	<b>279</b>
Introduction	279
Probability of $r$ Successes	279
Overdispersed Distribution	281
Uses of the Negative Binomial	282
References	283
<b>Part 5: Probability in Epidemiology and Medical Diagnosis</b>	<b>285</b>
<b>20. Some Epidemiological Considerations: Odds Ratio, Relative Risk, and Attributable Risk</b>	<b>287</b>
Basic Concepts	287
Introduction	287
Confidence Limits	291
Sample Size and Power	293
Attributable Risk (AR)	296
Relative Risks Below 1; Number Needed to Treat (NNT)	298
Advanced Concepts	300
Confidence Limits for Attributable Risk	300
Confidence Limits for NNT	302
Cautionary Tales	303
References	305

<b>21. Probability, Bayes' Theorem, Medical Diagnostic Evaluation, and Screening</b>	<b>307</b>
Bayes' Theorem Applied	307
Cautionary Tales	309
Sensitivity and Specificity	310
Cautionary Tales	315
Likelihood Ratios	319
Cutting Points	323
Receiver Operating Characteristic Curves	325
Some Comments on Screening Tests	330
References	331
 <b>Part 6: Comparing Means</b>	 <b>335</b>
<b>22. Comparison of Two Groups: t-Tests and Nonparametric Tests</b>	<b>337</b>
Basic Concepts	337
Introduction	337
Paired t-Test	338
Sample Size for Paired Test	346
Unpaired t-Test	346
Unequal Variances	349
Sample Size for Unpaired Test	350
Conclusion	350
Nonparametric or Distribution Free Tests	351
Advanced Concepts	356
Comparing Two Coefficients of Variation	356
The Paired t-Test Implies an Additive Model	356
Confidence Limits for Medians	358
Ranking Transforms	359
The Meaning of the Mann–Whitney Test	359
Appendix	360
References	361
 <b>23. t-Test Variants: Crossover Tests, Equivalence Tests</b>	 <b>363</b>
Crossover Trials	363
N of 1 trials	368
Equivalence and Noninferiority Testing	368
References	370
 <b>24. Multiple Comparisons</b>	 <b>373</b>
Introduction	373
Bonferroni Correction and Equivalent Tests	375
Error Rates	377



Group Sequential Boundaries	381
Sequential Analysis	384
Cautionary Tales	385
Adaptive Methods	386
References	387
<b>25. Analysis of Variance I. One-Way</b>	<b>391</b>
Basic Concepts	391
Basic Test	391
Homogeneity of Variance	399
Kruskal–Wallis Test	401
Independence of Observations	403
Effect Size	404
Sample Size	405
Advanced Concepts	406
Multiple Comparisons	406
Planned Experiments	417
References	419
<b>26. Analysis of Variance II. More Complex Forms</b>	<b>421</b>
Basic Concepts	421
Introduction	421
Two-Way ANOVA	423
Additivity	426
Multiple Factors	428
Friedman Test	429
Cochrane's Q Test	430
Interaction	431
Advanced and Alternative Concepts	436
Missing Data or Unequal Cell Sizes	436
Nested Designs	437
Transformations	442
Model II ANOVA	442
Repeated Measures Designs	444
Appendix	445
References	447
<b>Part 7: Regression and Correlation</b>	<b>449</b>
<b>27. Linear Regression</b>	<b>451</b>
Basic Concepts	451
Introduction	451
Exploratory Data Analysis	453

Transforming Curves	453
Model I Linear Regression Basics	455
Additional Calculations	466
Residuals	467
Confidence Limits	472
Comparison Methods (Basic)	475
Cautionary Tales	477
Advanced or Alternative Concepts	481
Heteroscedasticity	481
The Comparison Problem (Advanced)	483
Comparing Two or More Lines	486
Outliers	491
Ratio Measurements and Scaling Factors	496
Appendix	498
References	499

## **28. Variations Based on Linear Regression** **501**

Transforming the Y Variate	501
Inverse Prediction	502
Line of Best Fit Passes through Zero	504
Errors in the X Variate	505
Break Points	506
Resistant Lines	508
Appendix	509
Derivation of Formulas for a Straight Line Passing through Zero	509
References	511

## **29. Correlation** **513**

Basic Concepts	513
Introduction	513
Significance and Confidence Limits	515
Sample Size and Power	517
Cautionary Tales	519
Ordinal Numbers	524
Advanced and Alternative Concepts	527
Partial Correlation	527
Spurious Correlation	528
Cautionary Tales	530
Ratios and Scaling Factors	531
Intraclass Correlation	532
Appendix	534
References	535

<b>30. Multiple Regression</b>	<b>537</b>
Basic Concepts	537
Introduction	537
Multiple Linear Regression with Two $X$ Variables	538
Multiple Collinearity	540
Prerequisites	540
Determining the “Best” Regression	540
Nonlinear Regression: Polynomial Regression	544
Conclusions about Regression Methods in General	544
Advanced Concepts and Examples	546
Multiple Regression with Many Independent $X$ Variates	546
Detecting Multicollinearity	550
Determining the Correct Model	554
Nonlinear Regression: Polynomial Regression	565
Correcting for Multicollinearity	568
Principles of Nonlinear Regression	571
Dummy or Indicator Variables	571
Repeated Measures Regression	574
Multivariate Analysis	575
Longitudinal Regression	576
References	577
 <b>31. Serial Measurements: Time Series, Control Charts, Cusums</b>	 <b>579</b>
Introduction	579
Serial Correlation	580
Wald—Wolfowitz Runs Test	580
Up-and-Down Runs Test	581
Rank von Neumann Ratio	582
Ratio Measurements	584
Control Charts	585
Cumulative Sum Techniques (Cusums)	588
Serial Measurements	591
References	593
 <b>32. Dose—Response Analysis</b>	 <b>595</b>
General Principles	595
Quantal Dose—Response Curves	597
References	599

<b>33. Logistic Regression</b>	<b>601</b>
Introduction	601
Single Explanatory Variable	603
Multiple Explanatory Variables	606
Appropriateness of Model	610
References	611
<b>34. Poisson Regression</b>	<b>613</b>
Introduction	613
Suitability of Poisson Regression	614
Detecting Overdispersion	616
Correcting for Overdispersion	616
References	617
<b>Part 8: Miscellaneous Topics</b>	<b>619</b>
<b>35. Survival Analysis</b>	<b>621</b>
Basic Concepts	621
Introduction	621
Basic Method	623
Confidence Limits	628
Comparison of Different Survival Curves	629
Sample Size	630
Advanced Concepts	631
Calculating the Log-Rank Test	631
The Hazard Function	632
Hazard Ratio	634
Cox Proportional Hazards Regression	636
Competing Risks Analysis	640
References	642
<b>36. Meta-analysis</b>	<b>645</b>
Introduction	645
Forest Graphs	646
Funnel Plots	647
Radial Plots	649
L'Abbé Plots	650
Criticisms of Meta-analysis	651
References	651

<b>37. Resampling Statistics</b>	<b>655</b>
Introduction	655
Bootstrap	655
Permutations	657
Jackknife	659
Monte Carlo Methods	659
References	660
 <b>38. Study Design: Sampling, Clinical Trials</b>	 <b>663</b>
Sampling Problems	663
Historical Controls	666
Randomization	667
Clinical Trials	670
Placebo Effect	675
Alternatives to Randomized Clinical Trials	676
Propensity Analysis	676
References	677
 <b>Part 9: End Texts</b>	 <b>681</b>
<i>Answers to Problems</i>	683
<i>Glossary</i>	723
<i>Index</i>	727