Statistical Hypothesis Testing in Context

Reproducibility, Inference, and Science

Michael P. Fay
National Institute of Allergy and Infectious Diseases

Erica H. Brittain
National Institute of Allergy and Infectious Diseases

CAMBRIDGE
UNIVERSITY PRESS

Contents

Preface		*page* xi
1	**Introduction**	1
1.1	A Cautionary Tale	1
1.2	Science, Reproducibility, and Statistical Inferences	3
1.3	Using the Book	5
1.4	Notes	6
1.5	Problems	6
2	**Theory of Tests, *p*-Values, and Confidence Intervals**	8
2.1	Overview	8
2.2	Components of a Hypothesis Test	8
2.3	Confidence Sets: Inverting a Series of Hypothesis Tests	13
2.4	Properties of Hypothesis Tests	17
2.5	Summary	19
2.6	Extensions and Bibliographic Notes	20
2.7	Notes	20
2.8	Problems	22
3	**From Scientific Theory to Statistical Hypothesis Test**	23
3.1	The Statistician's Role	23
3.2	Causality versus Association	23
3.3	Designing Any Study	28
3.4	Designing an Experimental Study	33
3.5	Designing an Observational Study	36
3.6	A Probability Model for the Study	40
3.7	Statistics Methods Overview	43
3.8	Summary	45
3.9	Extensions and Bibliographic Notes	46
3.10	Notes	47
3.11	Problems	48
4	**One-Sample Studies with Binary Responses**	49
4.1	Overview	49
4.2	Key Method: Exact Central Method (Clopper–Pearson)	50
4.3	Other Methods for Independent Binary Responses	51
4.4	Designing a One-Sample Binary Response Study	60
4.5	Summary	62

4.6	Extensions and Bibliographic Notes	63
4.7	R Functions and Packages	64
4.8	Notes	65
4.9	Problems	66
5	**One-Sample Studies with Ordinal or Numeric Responses**	67
5.1	Overview	67
5.2	Exact Test on the Median or Other Quantiles	69
5.3	Tests on the Mean	70
5.4	Testing Other Parameters from One-Sample Numeric Data	77
5.5	Designing a One-Sample Study	79
5.6	Summary	79
5.7	Extensions and Bibliographic Notes	80
5.8	R Functions and Packages	80
5.9	Notes	80
5.10	Problems	83
6	**Paired Data**	84
6.1	Overview	84
6.2	Sign Test	85
6.3	Wilcoxon Signed-Rank Test	86
6.4	Choosing a Paired-Difference Test under Symmetry	90
6.5	Correlation and Related Types of Association	94
6.6	Agreement Coefficients	98
6.7	Summary	100
6.8	Extensions and Bibliographic Notes	101
6.9	R Functions and Packages	101
6.10	Notes	102
6.11	Problems	103
7	**Two-Sample Studies with Binary Responses**	104
7.1	Overview	104
7.2	Parameters to Measure Effects	106
7.3	Conditional Tests	108
7.4	Unconditional Tests	111
7.5	Mid-p Methods	115
7.6	Asymptotic Approximations	116
7.7	Comparison of *Binomial* Methods	117
7.8	Other Study Designs	119
7.9	Summary	120
7.10	Extensions and Bibliographic Notes	121
7.11	R Functions and Packages	122
7.12	Notes	122
7.13	Problems	122
8	**Assumptions and Hypothesis Tests**	124
8.1	Overview	124
8.2	Which Test Is More Restrictive?	124

8.3 Multiple Perspective Decision Rules 126
8.4 Summary 128
8.5 Extensions and Bibliographic Notes 128

9 Two-Sample Studies with Ordinal or Numeric Responses 129
9.1 Overview 129
9.2 Inferences on Proportional Odds 133
9.3 Differences in Means 133
9.4 Differences in Medians 141
9.5 Mann–Whitney Parameter 143
9.6 Choosing a Test when Power Concerns Dominate 151
9.7 Summary 155
9.8 Extensions and Bibliographic Notes 157
9.9 R Functions and Packages 157
9.10 Notes 158
9.11 Problems 158

10 General Methods for Frequentist Inferences 161
10.1 Overview 161
10.2 One-Dimensional Parameter 161
10.3 Pivots 164
10.4 Multivariate Normal Inferences 165
10.5 Asymptotic Normality and the Central Limit Theorem 166
10.6 Delta Method 168
10.7 Wald, Score, and Likelihood Ratio Tests 169
10.8 Sandwich Methods 174
10.9 Permutation Tests 176
10.10 Bootstrap 181
10.11 Melding: Combining Confidence Intervals in the Two-Sample Case 185
10.12 Within-Cluster Resampling 187
10.13 Other Methods 188
10.14 Summary 188
10.15 Extensions and Bibliographic Notes 190
10.16 R Functions and Packages 190
10.17 Notes 190
10.18 Problems 193

11 k-Sample Studies and Trend Tests 195
11.1 Overview 195
11.2 Unordered Groups: k-Sample Tests 196
11.3 Ordered Groups: Trend Tests 203
11.4 Follow-Up Tests after Independence Test and Familywise Error Rates 204
11.5 All Pairwise Comparisons 205
11.6 Many-to-One Comparisons 208
11.7 Max-t-Type Procedures 209
11.8 Summary 210
11.9 Extensions and Bibliographic Notes 211
11.10 R Functions and Packages 212
11.11 Notes 212
11.12 Problems 214

12	**Clustering and Stratification**	215
12.1	Overview	215
12.2	Clustering	215
12.3	Stratification	221
12.4	Summary	230
12.5	Extensions and Bibliographic Notes	232
12.6	R Functions and Packages	232
12.7	Notes	232
12.8	Problems	233
13	**Multiplicity in Testing**	235
13.1	Overview	235
13.2	Defining the Family of Hypothesis Tests	236
13.3	Properties of Multiple Testing Procedures	239
13.4	Procedures Using p-Values Only	240
13.5	Max-t-Type (Generalized Studentized Range) Inferences	241
13.6	Permutation- or Bootstrap-Based Methods	244
13.7	Sequential Testing of Different Hypotheses	246
13.8	Logical Constraints	247
13.9	Closure Method	248
13.10	Summary	248
13.11	Extensions and Bibliographic Notes	250
13.12	R Functions and Packages	250
13.13	Notes	250
13.14	Problems	251
14	**Testing from Models**	253
14.1	Overview	253
14.2	Linear Models	253
14.3	Generalized Linear Models	255
14.4	Ordinal Responses or Ranked Responses	259
14.5	Clustered Data	261
14.6	Covariate Selection	266
14.7	Some Problems When Testing from Models	268
14.8	Inference after Model Building	271
14.9	Summary	271
14.10	Extensions and Bibliographic Notes	273
14.11	R Functions and Packages	273
14.12	Notes	274
14.13	Problems	275
15	**Causality**	277
15.1	Overview	277
15.2	Causal Estimands	278
15.3	Why Causal Inference Is Difficult for Observational Studies	281
15.4	Causality and Experimental Studies	283
15.5	Causality and Interference	289
15.6	Causality and Observational Studies: Overview	290

15.7 Observational Studies: Adjustments When Confounders Are Measured 290
15.8 Observational Studies: Analysis with Unmeasured Confounders Using Instrumental Variables 297
15.9 Summary 298
15.10 Extensions and Bibliographic Notes 300
15.11 Notes 300
15.12 Problems 301

16 Censoring 302
16.1 Overview: Types of Censoring 302
16.2 One Sample with Right Censoring 303
16.3 Hazard Function 308
16.4 Two Samples with Right Censoring 309
16.5 Interval Censoring 318
16.6 Truncation 321
16.7 Summary 322
16.8 Extensions and Bibliographic Notes 323
16.9 R Functions and Packages 324
16.10 Notes 324
16.11 Problems 325

17 Missing Data 327
17.1 Overview 327
17.2 Missing Data Terminology 328
17.3 Sensitivity Analyses for Data Missing Not at Random 329
17.4 Methods for Missing at Random Data 333
17.5 Multiple Imputation 338
17.6 Summary 339
17.7 Extensions and Bibliographic Notes 340
17.8 R Functions and Packages 340
17.9 Problems 340

18 Group Sequential and Related Adaptive Methods 342
18.1 Overview 342
18.2 k Preplanned Interim Analyses 343
18.3 Arbitrary Number of Interim Analyses 346
18.4 When the Brownian Motion Approximation Does Not Hold 348
18.5 Estimates and Confidence Intervals after Interim Analyses 349
18.6 Two-Stage Adaptive Designs 352
18.7 Summary 354
18.8 Extensions and Bibliographic Notes 356
18.9 R Functions and Packages 356
18.10 Notes 356
18.11 Problems 357

19 Testing Fit, Equivalence, and Noninferiority 359
19.1 Overview 359
19.2 Testing Distributions and Lack of Fit 361

19.3 Equivalence Tests 365
19.4 Noninferiority Tests 366
19.5 Summary 372
19.6 Extensions and Bibliographic Notes 373
19.7 R Functions and Packages 373
19.8 Notes 373
19.9 Problems 375

20 Power and Sample Size 377
20.1 Overview 377
20.2 Power Calculations by Simulation 378
20.3 Power Calculations by Normal Approximations 380
20.4 Simulating the Sample Size 382
20.5 Nonadherence and Other Issues 383
20.6 Summary 385
20.7 Extensions and Bibliographic Notes 386
20.8 R Functions and Packages 386
20.9 Notes 386
20.10 Problems 388

21 Bayesian Hypothesis Testing 389
21.1 Introduction 389
21.2 Brief Overview of Bayesian Statistics 389
21.3 Bayesian Hypothesis Testing Using Bayes Factors 393
21.4 Bayesian Hypothesis Testing Using Decision Theory 394
21.5 When Can p-Values Be Interpreted in a Bayesian Manner? 396
21.6 Bibliographic Notes 401
21.7 Problems 402

References 404
Notation Index 420
Concept Index 424