

---

# *Contents*

---

<b>Preface</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Categorical Variables . . . . .	1
1.2 Contingency Tables . . . . .	2
1.3 Statistical Inference for Contingency Tables . . . . .	4
1.3.1 Notation and Concepts . . . . .	5
1.3.2 Effect Measures and Effect Estimates . . . . .	6
1.3.3 Confidence Intervals . . . . .	7
1.3.4 Hypothesis Tests . . . . .	7
1.4 Properties of Confidence Intervals and Hypothesis Tests . . . . .	8
1.4.1 Basic Properties . . . . .	8
1.4.2 Evaluation Criteria for Confidence Intervals . . . . .	10
1.4.3 Evaluation Criteria for Hypothesis Tests . . . . .	12
1.5 Generalized Linear Models . . . . .	13
1.5.1 Introduction . . . . .	13
1.5.2 Link Functions for the Binomial Distribution . . . . .	14
1.5.3 Loglinear Models versus Logit Link . . . . .	17
1.6 Exponential Family . . . . .	18
1.6.1 Definitions . . . . .	18
1.6.2 UMP Unbiased Tests . . . . .	19
1.7 Wald, Likelihood Ratio, and Score Inference . . . . .	23
1.8 Deviance, Likelihood Ratio Tests, and Pearson Chi-Squared Tests for Generalized Linear Models . . . . .	25
1.9 Exact Methods . . . . .	27
1.10 The Mid- $P$ Approach . . . . .	28
1.11 Bayes Inference . . . . .	29
<b>2 The <math>1 \times 2</math> Table and the Binomial Distribution</b>	<b>31</b>
2.1 Introduction . . . . .	31
2.2 Examples . . . . .	32
2.2.1 The Probability of Male Births to Indian Immigrants in Norway . . . . .	32

2.2.2	A Randomized Cross-Over Trial of Probiotics Versus Placebo for Patients with Irritable Bowel Syndrome . . . . .	33
2.3	Estimation of the Binomial Parameter $\pi$ . . . . .	33
2.4	Confidence Intervals for the Binomial Parameter . . . . .	34
2.4.1	The Wald Interval . . . . .	34
2.4.2	The Likelihood Ratio Interval . . . . .	35
2.4.3	The Wilson Score Interval . . . . .	35
2.4.4	The Agresti-Coull Interval . . . . .	36
2.4.5	The Jeffreys Interval . . . . .	37
2.4.6	The Arcsine Interval . . . . .	37
2.4.7	Exact Intervals . . . . .	38
2.4.8	Mid- $P$ Intervals . . . . .	39
2.4.9	Examples . . . . .	40
2.4.10	Evaluation of Intervals . . . . .	42
2.5	Tests for the Binomial Parameter . . . . .	51
2.5.1	The Null and Alternative Hypotheses . . . . .	51
2.5.2	The Wald Test . . . . .	51
2.5.3	The Likelihood Ratio Test . . . . .	52
2.5.4	The Score Test . . . . .	52
2.5.5	Exact Tests . . . . .	53
2.5.6	Mid- $P$ Tests . . . . .	53
2.5.7	Examples . . . . .	54
2.5.8	Evaluation of Tests . . . . .	56
2.6	Recommendations . . . . .	62
2.6.1	Summary . . . . .	62
2.6.2	Confidence Intervals for the Binomial Parameter . . . . .	63
2.6.3	Tests for the Binomial Parameter . . . . .	65
<b>3</b>	<b>The <math>1 \times c</math> Table and the Multinomial Distribution</b> . . . . .	<b>67</b>
3.1	Introduction . . . . .	67
3.2	Example: Distribution of Genotype Counts of SNP rs6498169 in Patients with Rheumatoid Arthritis . . . . .	68
3.3	Tests for the Multinomial Parameters . . . . .	68
3.3.1	Estimation of the Multinomial Parameters . . . . .	68
3.3.2	The Null and Alternative Hypotheses . . . . .	69
3.3.3	The Pearson Chi-Squared Test . . . . .	69
3.3.4	The Likelihood Ratio Test . . . . .	70
3.3.5	The Exact Multinomial Test . . . . .	70
3.3.6	The Mid- $P$ Multinomial Test . . . . .	71
3.3.7	Example . . . . .	71
3.3.8	Evaluation of Tests . . . . .	73
3.4	An Incompletely Specified Null Hypothesis . . . . .	76
3.5	Ordered-Restricted Trend Tests . . . . .	78
3.6	Confidence Intervals for the Multinomial Parameters . . . . .	79

3.6.1	Simultaneous Confidence Intervals for the Multinomial Parameters . . . . .	80
3.6.2	Example . . . . .	82
3.7	Recommendations . . . . .	83
3.7.1	Summary . . . . .	83
3.7.2	Tests for the Multinomial Parameters . . . . .	84
3.7.3	Confidence Intervals for the Multinomial Parameters . . . . .	84
<b>4</b>	<b>The <math>2 \times 2</math> Table</b>	<b>85</b>
4.1	Introduction . . . . .	85
4.2	Sampling Models, Study Designs, and Examples . . . . .	86
4.2.1	Both Margins Fixed Model . . . . .	86
4.2.2	Row Margins Fixed Model . . . . .	86
4.2.3	Column Margins Fixed Model . . . . .	87
4.2.4	Total Sum Fixed Model . . . . .	87
4.2.5	No Sums Fixed Model . . . . .	88
4.2.6	Inverse Sampling . . . . .	89
4.3	Probabilities, Probability Models, and Link Functions . . . . .	89
4.3.1	Probabilities and Probability Models . . . . .	89
4.3.2	Link Functions for the $2 \times 2$ Table . . . . .	92
4.4	Tests for Association . . . . .	93
4.4.1	The Null and Alternative Hypotheses . . . . .	93
4.4.2	The Pearson Chi-Squared Test . . . . .	93
4.4.3	The Likelihood Ratio Test . . . . .	94
4.4.4	Tests Based on the Difference between Proportions . . . . .	95
4.4.5	The Fisher Exact Test . . . . .	96
4.4.6	The Fisher Mid- $P$ Test . . . . .	97
4.4.7	Exact Unconditional Tests . . . . .	98
4.4.8	Examples . . . . .	99
4.4.9	Evaluation of Tests . . . . .	105
4.5	Confidence Intervals for the Difference between Probabilities . . . . .	113
4.5.1	Introduction and Estimation . . . . .	113
4.5.2	The Wald Interval . . . . .	115
4.5.3	The Agresti-Caffo Interval . . . . .	115
4.5.4	The Newcombe Hybrid Score Interval . . . . .	116
4.5.5	Asymptotic Score Intervals . . . . .	116
4.5.6	Exact Unconditional Intervals . . . . .	117
4.5.7	Examples . . . . .	119
4.5.8	Evaluation of Intervals . . . . .	123
4.6	Confidence Intervals for the Number Needed to Treat . . . . .	130
4.6.1	Introduction . . . . .	130
4.6.2	Estimation . . . . .	130
4.6.3	Confidence Intervals . . . . .	131
4.6.4	Examples . . . . .	131

4.7	Confidence Intervals for the Ratio of Probabilities . . . . .	133
4.7.1	Introduction and Estimation . . . . .	133
4.7.2	Log Intervals . . . . .	134
4.7.3	The Price and Bonett Approximate Bayes Interval . . . . .	135
4.7.4	Inverse Hyperbolic Sine Intervals . . . . .	136
4.7.5	The MOVER-R Wilson Interval . . . . .	137
4.7.6	Asymptotic Score Intervals . . . . .	137
4.7.7	Exact Unconditional Intervals . . . . .	139
4.7.8	Examples . . . . .	141
4.7.9	Evaluation of Intervals . . . . .	145
4.8	Confidence Intervals for the Odds Ratio . . . . .	151
4.8.1	Introduction and Estimation . . . . .	151
4.8.2	Logit Intervals . . . . .	153
4.8.3	Inverse Hyperbolic Sine Intervals . . . . .	154
4.8.4	The MOVER-R Wilson Interval . . . . .	155
4.8.5	Asymptotic Score Intervals . . . . .	155
4.8.6	Exact Conditional Intervals . . . . .	157
4.8.7	Mid- $P$ Intervals . . . . .	158
4.8.8	Exact Unconditional Intervals . . . . .	158
4.8.9	Examples . . . . .	160
4.8.10	Evaluation of Intervals . . . . .	166
4.9	Recommendations . . . . .	172
4.9.1	Summary . . . . .	172
4.9.2	Tests for Association . . . . .	174
4.9.3	Confidence Intervals for the Difference between Probabilities . . . . .	175
4.9.4	Confidence Intervals for the Number Needed to Treat . . . . .	176
4.9.5	Confidence Intervals for the Ratio of Probabilities . . . . .	176
4.9.6	Confidence Intervals for the Odds Ratio . . . . .	177
<b>5</b>	<b>The Ordered <math>r \times 2</math> Table</b>	<b>179</b>
5.1	Introduction . . . . .	179
5.2	Examples . . . . .	180
5.2.1	Alcohol Consumption and Malformations . . . . .	180
5.2.2	Elevated Troponin T Levels in Stroke Patients . . . . .	181
5.3	Notation and Statistical Models . . . . .	181
5.3.1	Models When We Have No Information on the Grouping . . . . .	182
5.3.2	Models with Scores Assigned to the Groups . . . . .	183
5.3.3	Trend Analysis in Case-Control Studies . . . . .	183
5.3.4	Choice of Link Function . . . . .	184
5.4	Designs and Sampling Distributions . . . . .	185
5.5	Assigning Scores . . . . .	187
5.6	Tests for Unspecific Ordering . . . . .	187
5.6.1	The Null and Alternative Hypotheses . . . . .	187

5.6.2	The Pearson Chi-Squared Test for Unspecific Ordering	188
5.6.3	The Likelihood Ratio Test for Unspecific Ordering	189
5.6.4	Exact and Mid- $P$ Tests for Unspecific Ordering	189
5.6.5	Examples	190
5.6.6	Evaluation of Tests	192
5.7	Tests for Trend in the Linear and Logit Models	195
5.7.1	The Null and Alternative Hypotheses	195
5.7.2	The Cochran-Armitage Test	196
5.7.3	The Modified Cochran-Armitage Test	198
5.7.4	The Cochran-Armitage Exact Conditional Test	198
5.7.5	The Cochran-Armitage Mid- $P$ Test	199
5.7.6	The Mantel-Haenszel Test for Trend	199
5.7.7	The Wald Test for the Linear Model	200
5.7.8	The Likelihood Ratio Test for the Linear Model	201
5.7.9	The Wald Test for the Logit Model	201
5.7.10	The Likelihood Ratio Test for the Logit Model	201
5.7.11	The Exact Conditional and Mid- $P$ Tests for the Logit Model	202
5.7.12	An Exact Unconditional Test	202
5.7.13	Testing the Fit of a Model with Linear or Logit Trend	203
5.7.14	Examples	204
5.7.15	Evaluation of Tests for Trend	206
5.8	Confidence Intervals for the Trend	211
5.8.1	The Cochran-Armitage Interval	211
5.8.2	Wald Intervals	212
5.8.3	The Profile Likelihood Interval	212
5.8.4	The Exact Conditional Interval	213
5.8.5	The Mid- $P$ Interval	213
5.8.6	Examples	214
5.8.7	Evaluation of Intervals	215
5.9	Recommendations	219
5.9.1	Summary	219
5.9.2	Tests for Unspecific Ordering	220
5.9.3	Tests for Trend	221
5.9.4	Confidence Intervals for the Trend	221
<b>6</b>	<b>The Ordered <math>2 \times c</math> Table</b>	<b>223</b>
6.1	Introduction	223
6.2	Examples	224
6.2.1	The Adolescent Placement Study	224
6.2.2	Postoperative Nausea	225
6.3	Notation and Sampling Distributions	225
6.4	Accounting for the Ordering with Odds Ratios	227
6.5	Local Odds Ratios	228

6.5.1	The Pearson Chi-Squared and Likelihood Ratio Tests for Local Odds Ratios . . . . .	229
6.5.2	Exact and Mid- $P$ Tests for Local Odds Ratios . . . . .	229
6.5.3	Example . . . . .	230
6.5.4	Evaluation of Tests . . . . .	232
6.6	Cumulative Odds Ratios . . . . .	232
6.6.1	The Pearson Chi-Squared and Likelihood Ratio Tests for Cumulative Odds Ratios . . . . .	232
6.6.2	Exact and Mid- $P$ Tests for Cumulative Odds Ratios . . . . .	233
6.6.3	Example . . . . .	234
6.6.4	Evaluation of Tests . . . . .	235
6.7	Continuation Odds Ratios . . . . .	236
6.8	Ranking of Observations . . . . .	236
6.9	Mantel-Haenszel Test of Association with Column Scores . . . . .	237
6.9.1	Example . . . . .	238
6.10	Logit Models for Ordinal Data . . . . .	239
6.10.1	Cumulative Logits . . . . .	239
6.10.2	Adjacent Category . . . . .	241
6.10.3	Continuation Ratio . . . . .	241
6.10.4	Example . . . . .	242
6.11	The Proportional Odds Model . . . . .	244
6.12	Estimating Proportional Odds Models . . . . .	245
6.13	Testing the Fit of a Proportional Odds Model . . . . .	247
6.13.1	The Pearson Goodness-of-Fit Test . . . . .	247
6.13.2	The Pearson Residuals . . . . .	248
6.13.3	The Likelihood Ratio (Deviance) Test . . . . .	248
6.13.4	The Brant Test . . . . .	248
6.14	Testing for Effect in a Proportional Odds Model . . . . .	250
6.14.1	The Wald Test . . . . .	250
6.14.2	The Likelihood Ratio Test . . . . .	251
6.14.3	The Score Test (the Wilcoxon-Mann-Whitney Test) . . . . .	251
6.14.4	The Exact Conditional Linear Rank Test . . . . .	252
6.14.5	The Mid- $P$ Linear Rank Test . . . . .	253
6.14.6	An Exact Unconditional Test . . . . .	253
6.15	The Cumulative Probit Model . . . . .	254
6.15.1	Estimating Cumulative Probit Models . . . . .	255
6.15.2	Testing the Fit of a Cumulative Probit Model . . . . .	257
6.15.3	Testing for Effect in a Cumulative Probit Model . . . . .	257
6.16	Examples of Testing for Fit and Effect in the Proportional Odds and Cumulative Probit Models . . . . .	259
6.17	Evaluation of Tests for Effect in the Proportional Odds and Cumulative Probit Models . . . . .	261
6.18	Confidence Intervals for the Proportional Odds and Cumulative Probit Models . . . . .	266
6.18.1	The Wald Interval . . . . .	266

6.18.2	The Profile Likelihood Interval . . . . .	267
6.18.3	Examples . . . . .	267
6.18.4	Evaluations of Intervals . . . . .	268
6.19	Recommendations . . . . .	271
6.19.1	Summary . . . . .	271
6.19.2	Testing the Fit of a Proportional Odds or Cumulative Probit Model . . . . .	272
6.19.3	Tests for Effect in the Proportional Odds Model . . .	273
6.19.4	Tests for Effect in the Cumulative Probit Model . . .	273
6.19.5	Confidence Intervals for the Proportional Odds and Cu- mulative Probit Models . . . . .	273
<b>7</b>	<b>The <math>r \times c</math> Table</b>	<b>275</b>
7.1	Introduction . . . . .	275
7.2	Examples . . . . .	277
7.2.1	Treatment for Ear Infection (Unordered $3 \times 2$ Table) .	277
7.2.2	Psychiatric Diagnoses and Physical Activity (Unordered $6 \times 2$ Table) . . . . .	277
7.2.3	Psychiatric Diagnoses and BMI (Unordered or Singly Ordered $6 \times 3$ Table) . . . . .	277
7.2.4	Low Birth Weight and Psychiatric Morbidity (Singly Ordered $3 \times 3$ Table) . . . . .	278
7.2.5	Colorectal Cancer (Doubly Ordered $4 \times 4$ Table) . . .	279
7.2.6	Breast Tumor (Doubly Ordered $3 \times 5$ Table) . . . . .	280
7.2.7	Self-Rated Health (Doubly Ordered $4 \times 4$ Table) . . .	280
7.3	Notation and Sampling Distributions . . . . .	281
7.4	Odds Ratios in $r \times c$ Tables . . . . .	283
7.5	Unordered $r \times c$ Tables . . . . .	285
7.5.1	The Pearson Chi-Squared Test . . . . .	285
7.5.2	The Likelihood Ratio Test . . . . .	286
7.5.3	The Fisher-Freeman-Halton and Related Exact Tests .	286
7.5.4	Mid- $P$ Tests . . . . .	287
7.5.5	The Fisher-Freeman-Halton Asymptotic Test . . . . .	287
7.5.6	Exact Unconditional Tests . . . . .	288
7.5.7	Identifying Cells with Significantly Deviating Counts .	288
7.5.8	Multiple Comparisons of Binomial Probabilities . . . .	289
7.5.9	Examples . . . . .	292
7.5.10	Evaluations of Tests . . . . .	296
7.6	Singly Ordered $r \times c$ Tables . . . . .	300
7.6.1	Introduction and Null Hypothesis . . . . .	300
7.6.2	The Kruskal-Wallis Asymptotic Test . . . . .	301
7.6.3	The Kruskal-Wallis Exact and Mid- $P$ Tests . . . . .	301
7.6.4	Examples . . . . .	302
7.6.5	Evaluations of Tests . . . . .	304

7.7	Ordinal Logistic Regression with a Nominal Covariate for Singly Ordered $r \times c$ Tables . . . . .	306
7.7.1	The Cumulative Logit Model . . . . .	306
7.7.2	Estimating Proportional Odds Models . . . . .	307
7.7.3	Testing the Fit of a Proportional Odds Model . . . . .	307
7.7.4	Testing for Effect in a Proportional Odds Model . . . . .	308
7.7.5	Confidence Intervals for the Effect in a Proportional Odds Model . . . . .	308
7.7.6	Examples . . . . .	309
7.8	Tests for Association in Doubly Ordered $r \times c$ Tables . . . . .	311
7.8.1	The Linear-by-Linear Test for Association . . . . .	311
7.8.2	The Jonckheere-Terpstra Test for Association . . . . .	312
7.8.3	Exact and Mid- $P$ Tests . . . . .	313
7.8.4	Examples . . . . .	314
7.8.5	Evaluations of Tests . . . . .	316
7.9	Correlation Measures in Doubly Ordered $r \times c$ Tables . . . . .	316
7.9.1	The Pearson Correlation Coefficient . . . . .	317
7.9.2	The Spearman Correlation Coefficient . . . . .	318
7.9.3	The Gamma Coefficient . . . . .	319
7.9.4	The Kendall Rank Correlation Coefficient . . . . .	320
7.9.5	Polychoric Correlation . . . . .	321
7.9.6	Examples . . . . .	322
7.10	Ordinal Logistic Regression with an Ordinal Covariate for Doubly Ordered $r \times c$ Tables . . . . .	324
7.10.1	The Cumulative Logit Model . . . . .	324
7.10.2	Estimating Proportional Odds Models . . . . .	324
7.10.3	Testing the Fit of a Proportional Odds Model . . . . .	325
7.10.4	Testing for Effect in a Proportional Odds Model . . . . .	325
7.10.5	Confidence Intervals for the Effect in a Proportional Odds Model . . . . .	326
7.10.6	Examples . . . . .	326
7.11	Recommendations . . . . .	326
7.11.1	Summary . . . . .	326
7.11.2	Unordered Tables . . . . .	327
7.11.3	Singly Ordered Tables . . . . .	328
7.11.4	Doubly Ordered Tables . . . . .	329
<b>8</b>	<b>The Paired <math>2 \times 2</math> Table</b>	<b>331</b>
8.1	Introduction . . . . .	331
8.2	Examples . . . . .	332
8.2.1	Airway Hyper-Responsiveness Status before and after Stem Cell Transplantation . . . . .	332
8.2.2	Complete Response before and after Consolidation Therapy . . . . .	333



8.2.3	The Association between Floppy Eyelid Syndrome and Obstructive Sleep Apnea-Hypopnea Syndrome . . . .	334
8.3	Notation and Sampling Distribution . . . . .	335
8.4	Statistical Models . . . . .	336
8.4.1	Marginal Models . . . . .	336
8.4.2	Subject-Specific Models . . . . .	336
8.5	Tests for Association . . . . .	338
8.5.1	The Null and Alternative Hypotheses . . . . .	338
8.5.2	The McNemar Asymptotic Test . . . . .	338
8.5.3	The McNemar Exact Conditional Test . . . . .	339
8.5.4	The McNemar Mid- $P$ Test . . . . .	340
8.5.5	The McNemar Exact Unconditional Test . . . . .	340
8.5.6	Examples . . . . .	341
8.5.7	Evaluation of Tests . . . . .	345
8.6	Confidence Intervals for the Difference between Probabilities . . . . .	350
8.6.1	Introduction and Estimation . . . . .	350
8.6.2	Wald Intervals . . . . .	351
8.6.3	The Newcombe Square-And-Add Interval (MOVER Wilson Score) . . . . .	352
8.6.4	The Tango Asymptotic Score Interval . . . . .	353
8.6.5	The Sidik Exact Unconditional Interval . . . . .	354
8.6.6	Examples . . . . .	355
8.6.7	Evaluation of Intervals . . . . .	356
8.7	Confidence Intervals for the Number Needed to Treat . . . . .	362
8.7.1	Introduction and Estimation . . . . .	362
8.7.2	Confidence Intervals . . . . .	363
8.7.3	Examples . . . . .	363
8.8	Confidence Intervals for the Ratio of Probabilities . . . . .	365
8.8.1	Introduction and Estimation . . . . .	365
8.8.2	The Wald Interval . . . . .	366
8.8.3	The Tang Asymptotic Score Interval . . . . .	366
8.8.4	The Bonett-Price Hybrid Wilson Score Interval . . . . .	366
8.8.5	The MOVER Wilson Score Interval . . . . .	368
8.8.6	Examples . . . . .	368
8.8.7	Evaluation of Intervals . . . . .	370
8.9	Confidence Intervals for the Odds Ratio . . . . .	373
8.9.1	Introduction and Estimation . . . . .	373
8.9.2	The Wald Interval . . . . .	375
8.9.3	The Wald Interval with Laplace Adjustment . . . . .	376
8.9.4	Intervals Obtained by Transforming Intervals for $\pi_{12}/(\pi_{12} + \pi_{21})$ . . . . .	376
8.9.5	Examples . . . . .	378
8.9.6	Evaluation of Intervals . . . . .	378
8.10	Recommendations . . . . .	382
8.10.1	Summary . . . . .	382

8.10.2	Tests for Association . . . . .	383
8.10.3	Confidence Intervals for the Difference between Probabilities . . . . .	384
8.10.4	Confidence Intervals for the Number Needed to Treat . . . . .	385
8.10.5	Confidence Intervals for the Ratio of Probabilities . . . . .	385
8.10.6	Confidence Intervals for the Odds Ratio . . . . .	386
<b>9</b>	<b>The Paired <math>c \times c</math> Table</b>	<b>387</b>
9.1	Introduction . . . . .	387
9.2	Examples . . . . .	388
9.2.1	Pretherapy Susceptibility of Pathogens in Patients with Complicated Urinary Tract Infection . . . . .	388
9.2.2	A Comparison between Serial and Retrospective Measurements of Change in Patients' Health Status . . . . .	388
9.3	Notation, Sampling Distribution, and Hypotheses . . . . .	390
9.4	Nominal Categories . . . . .	390
9.4.1	A Marginal Model . . . . .	391
9.4.2	A Subject-Specific Model . . . . .	391
9.4.3	The Bhapkar Test for Marginal Homogeneity . . . . .	391
9.4.4	The Stuart Test for Marginal Homogeneity . . . . .	392
9.4.5	Closed-Form Expression of the Stuart Test . . . . .	392
9.4.6	Tests and Confidence Intervals for Individual Categories . . . . .	393
9.4.7	The McNemar-Bowker Test for Symmetry . . . . .	395
9.4.8	Examples . . . . .	395
9.4.9	Evaluation of Tests . . . . .	397
9.4.10	Evaluation of Intervals . . . . .	402
9.5	Ordinal Categories . . . . .	404
9.5.1	A Marginal Model . . . . .	405
9.5.2	A Subject-Specific Model . . . . .	406
9.5.3	Simple Test for Three-Level Outcomes . . . . .	406
9.5.4	Wald Test and Interval for Marginal Mean Scores . . . . .	407
9.5.5	Score Test and Interval for Marginal Mean Scores . . . . .	408
9.5.6	Wald Test and Interval for Marginal Mean Ranks/Ridits . . . . .	409
9.5.7	Wald Logit Test and Interval for Marginal Mean Ranks/Ridits . . . . .	411
9.5.8	Examples . . . . .	411
9.5.9	Evaluation of Tests . . . . .	414
9.5.10	Evaluation of Intervals . . . . .	417
9.6	Recommendations . . . . .	422
9.6.1	Summary . . . . .	422
9.6.2	Nominal Categories . . . . .	422
9.6.3	Ordinal Categories . . . . .	423

<b>10 Stratified <math>2 \times 2</math> Tables and Meta-Analysis</b>	<b>425</b>
10.1 Introduction . . . . .	425
10.2 Examples . . . . .	427
10.2.1 Smoking and Lung Cancer . . . . .	427
10.2.2 Prophylactic Use of Lidocaine in Myocardial Infarction (Meta-Analysis) . . . . .	427
10.3 Notation and Sampling Designs . . . . .	428
10.4 Statistical Models, Effect Measures, and Effect Estimates . .	431
10.5 Sampling Distributions . . . . .	433
10.6 A Guide to the Analyses of Stratified $2 \times 2$ Tables . . . . .	435
10.7 Confidence Intervals for the Stratum Specific Effects . . . . .	437
10.8 Estimating a Common Effect . . . . .	438
10.8.1 Mantel-Haenszel and Inverse Variance Estimates . . .	439
10.8.2 The Peto Odds Ratio Estimate . . . . .	440
10.8.3 Maximum Likelihood Estimates . . . . .	441
10.9 Tests of Homogeneity of Effects over the $K$ Strata . . . . .	442
10.9.1 The Cochran $Q$ Test . . . . .	442
10.9.2 The Likelihood Ratio Test . . . . .	443
10.9.3 The Pearson Chi-Squared Test . . . . .	443
10.9.4 Evaluations of Tests . . . . .	443
10.10 Additional Tests of Homogeneity of Odds Ratios over the $K$ Strata . . . . .	446
10.10.1 The Breslow-Day Test . . . . .	449
10.10.2 Exact and Mid- $P$ Tests . . . . .	450
10.10.3 The Conditional Score Test . . . . .	452
10.10.4 The Peto Test . . . . .	453
10.10.5 Evaluations of Tests . . . . .	453
10.11 Tests of a Common Difference between Probabilities . . . . .	454
10.11.1 The Likelihood Ratio Test . . . . .	455
10.11.2 The Pearson Chi-Squared Test . . . . .	455
10.11.3 The Wald Test Based on the Mantel-Haenszel Estimate	456
10.11.4 The Wald Test Based on the Inverse Variance Estimate	457
10.11.5 Evaluations of Tests . . . . .	457
10.12 Tests of a Common Ratio of Probabilities . . . . .	457
10.12.1 The Likelihood Ratio Test . . . . .	459
10.12.2 The Pearson Chi-Squared Test . . . . .	459
10.12.3 The Wald Test Based on the Mantel-Haenszel Estimate	460
10.12.4 The Wald Test Based on the Inverse Variance Estimate	460
10.12.5 Evaluations of Tests . . . . .	461
10.13 Tests of a Common Odds Ratio . . . . .	461
10.13.1 The Likelihood Ratio Test . . . . .	463
10.13.2 The Pearson Chi-Squared Test . . . . .	463
10.13.3 The Cochran-Mantel-Haenszel Test . . . . .	464
10.13.4 The RBG Test . . . . .	465

10.13.5	The Woolf Test . . . . .	466
10.13.6	The Gart Exact, Mid- $P$ , and Asymptotic Tests . . . . .	467
10.13.7	An Exact Unconditional Test . . . . .	467
10.13.8	Evaluations of Tests . . . . .	468
10.14	Confidence Intervals for the Common Effect . . . . .	468
10.14.1	Asymptotic Confidence Intervals . . . . .	469
10.14.2	The Gart Exact and Mid- $P$ Intervals . . . . .	470
10.14.3	Evaluations of Intervals . . . . .	472
10.15	Random Effects Models for Stratified $2 \times 2$ Tables . . . . .	474
10.15.1	Introduction . . . . .	474
10.15.2	The DerSimonian-Laird Estimate and Confidence Interval . . . . .	478
10.15.3	The Restricted Maximum Likelihood Estimate and Confidence Interval . . . . .	479
10.16	Examples Analyzed . . . . .	480
10.16.1	Smoking and Lung Cancer . . . . .	480
10.16.2	Prophylactic Use of Lidocaine in Myocardial Infarction (Meta-Analysis) . . . . .	480
10.17	Additional Methods for Meta-Analysis . . . . .	483
10.18	Recommendations . . . . .	489
10.18.1	Testing Homogeneity of Effects over the $K$ Strata . . . . .	489
10.18.2	Tests of a Common Effect . . . . .	491
10.18.3	Estimates and Confidence Intervals . . . . .	491
<b>11</b>	<b>Other Stratified Tables</b>	<b>493</b>
11.1	Introduction . . . . .	493
11.2	Examples . . . . .	493
11.2.1	Los Angeles Retirement Community Study of Endometrial Cancer (Stratified Paired $2 \times 2$ Tables) . . . . .	493
11.2.2	Ille-et-Vilaine Study of Oesophageal Cancer (Stratified $r \times 2$ Tables) . . . . .	494
11.2.3	Adolescent Placement Study (Stratified $2 \times c$ Tables) . . . . .	494
11.2.4	Psychiatric Diagnoses and BMI (Stratified $r \times c$ Tables) . . . . .	495
11.3	Stratified Paired $2 \times 2$ Tables . . . . .	497
11.3.1	Notation and Statistical Model . . . . .	497
11.3.2	Testing Homogeneity over the Strata . . . . .	498
11.3.3	Estimation and Testing of a Common Effect over the Strata . . . . .	499
11.3.4	Example . . . . .	500
11.4	Stratified $r \times 2$ Tables . . . . .	501
11.4.1	Introduction . . . . .	501
11.4.2	Notation and Statistical Model . . . . .	501
11.4.3	The Likelihood Ratio Test of Homogeneity of Trend over the $K$ Strata . . . . .	502

11.4.4	The Wald Test of a Common Trend . . . . .	502
11.4.5	The Cochran-Mantel-Haenszel Asymptotic, Exact, and Mid- <i>P</i> Tests of a Common Trend . . . . .	503
11.4.6	Example . . . . .	503
11.5	Stratified $2 \times c$ Tables . . . . .	504
11.5.1	Unordered Categories . . . . .	504
11.5.2	Ordered Categories . . . . .	506
11.5.3	The Proportional Odds Model for Stratified $2 \times c$ Tables	506
11.5.4	Example . . . . .	507
11.6	Stratified $r \times c$ Tables . . . . .	508
11.6.1	Stratified Unordered Tables . . . . .	509
11.6.2	Stratified Singly Ordered Tables . . . . .	510
11.6.3	Stratified Doubly Ordered Tables . . . . .	511
11.6.4	Example . . . . .	512
11.7	Recommendations . . . . .	513
11.7.1	Stratified Paired $2 \times 2$ Tables . . . . .	513
11.7.2	Stratified $r \times 2$ , $2 \times c$ , and $r \times c$ Tables . . . . .	513
<b>12 Sample Size Calculations</b>		<b>515</b>
12.1	Introduction . . . . .	515
12.2	General Principles . . . . .	516
12.2.1	Testing for No Effect . . . . .	517
12.2.2	Testing for Equivalence and Non-inferiority . . . . .	518
12.2.3	Sample Size Calculations based on Confidence Interval Estimation . . . . .	521
12.3	The $1 \times 2$ Table . . . . .	521
12.3.1	The Score Test . . . . .	522
12.3.2	Exact Binomial Test . . . . .	522
12.3.3	Example . . . . .	523
12.4	The $2 \times 2$ Table . . . . .	524
12.4.1	The Score Test (Pearson Chi-Squared Test) . . . . .	524
12.4.2	Exact Conditional Test . . . . .	525
12.4.3	Confidence Interval Method . . . . .	525
12.4.4	Sample Size for Equivalence and Non-Inferiority . . . . .	526
12.4.5	Example . . . . .	526
12.5	The Ordered $r \times 2$ Table . . . . .	527
12.5.1	Sample Size Requirements for Ordered $r \times 2$ Tables . . . . .	528
12.5.2	Example . . . . .	529
12.6	The Ordered $2 \times c$ Table . . . . .	529
12.7	The Unordered $r \times c$ Table . . . . .	530
12.7.1	Sample Size Requirements for Unordered $r \times c$ Tables . . . . .	531
12.7.2	Example . . . . .	532
12.8	Singly and Doubly Ordered $r \times c$ Tables . . . . .	532
12.9	The Paired $2 \times 2$ Table . . . . .	533

12.9.1	Sample Size Estimation when Testing for Marginal Homogeneity . . . . .	533
12.9.2	The McNemar Asymptotic Test . . . . .	533
12.9.3	The McNemar Exact Conditional Test . . . . .	534
12.9.4	The McNemar Exact Unconditional Test . . . . .	534
12.9.5	Sample Size Calculations for Equivalence and Non-Inferiority . . . . .	534
12.9.6	Example . . . . .	535
12.10	The Paired $c \times c$ Table . . . . .	536
12.10.1	Sample Size Calculations for Testing Symmetry . . . . .	536
12.11	Stratified $2 \times 2$ Tables . . . . .	537
12.11.1	Sample Size Requirements for Testing for Common Effect of the Odds Ratio . . . . .	537
12.11.2	Example . . . . .	538
12.12	Recommendations . . . . .	538
<b>13</b>	<b>Miscellaneous Topics</b>	<b>541</b>
13.1	Diagnostic Accuracy . . . . .	541
13.1.1	Introduction . . . . .	541
13.1.2	Measures of Diagnostic Accuracy . . . . .	541
13.1.3	Confidence Intervals and Hypothesis Tests for a Single Diagnostic Test . . . . .	545
13.1.4	Example: Chlamydia Rapid Test versus Polymerase Chain Reaction . . . . .	546
13.1.5	Confidence Intervals and Hypothesis Tests for Two Diagnostic Tests . . . . .	547
13.2	Inter-Rater Agreement . . . . .	548
13.2.1	Introduction . . . . .	548
13.2.2	Cohen's Kappa for Dichotomous or Nominal Categories . . . . .	549
13.2.3	Cohen's Weighted Kappa . . . . .	551
13.2.4	Other Measures of Agreement . . . . .	552
13.2.5	More Than Two Raters . . . . .	552
13.2.6	The Kendall Coefficient of Concordance . . . . .	552
13.2.7	Positive and Negative Agreement for $2 \times 2$ Tables . . . . .	553
13.2.8	Example: Children with Cerebral Palsy . . . . .	553
13.2.9	Recommendations . . . . .	555
13.3	Missing Data . . . . .	556
13.3.1	Introduction . . . . .	556
13.3.2	Missing Data Mechanisms: MCAR, MAR, and MNAR . . . . .	558
13.3.3	Estimation in a $2 \times 2$ Table . . . . .	559
13.3.4	Estimation in Stratified $2 \times 2$ Tables . . . . .	561
13.3.5	Hypothesis Tests . . . . .	561
13.3.6	Methods for MNAR . . . . .	562
13.4	Structural Zeros . . . . .	562

13.4.1	Introduction . . . . .	562
13.4.2	Structural Zero Due to Design . . . . .	563
13.4.3	Quasi-Independence . . . . .	566
13.5	Categorization of Continuous Variables . . . . .	568
13.5.1	Arguments Used for Categorization . . . . .	569
13.5.2	How to Categorize . . . . .	571
13.5.3	How to Assign Scores . . . . .	571
13.5.4	Recommendations . . . . .	572
13.6	Ecological Inference . . . . .	572
13.6.1	Introduction . . . . .	572
13.6.2	Ecological Inference for Stratified $2 \times 2$ Tables . . . . .	573
13.6.3	Method of Bounds . . . . .	574
13.6.4	Ecological Regression . . . . .	574
13.6.5	Maximum Likelihood . . . . .	575
13.6.6	Hierarchical Models . . . . .	575
<b>A</b>	<b>Appendix</b>	<b>577</b>
A.1	List of Statistical Methods . . . . .	577
A.2	List of Examples . . . . .	587
	<b>Bibliography</b>	<b>591</b>
	<b>Index</b>	<b>619</b>