

Contents

Preface to the second edition	ix
Preface to the first edition	xiii
The Authors	xvii
1 Analyzing a binary response, part 1: introduction	1
1.1 One binary variable	1
1.1.1 Bernoulli and binomial probability distributions	1
1.1.2 Inference for the probability of success	8
1.1.3 True confidence levels for confidence intervals	18
1.2 Two binary variables	24
1.2.1 Notation and model	26
1.2.2 Confidence intervals for the difference of two probabilities	30
1.2.3 Test for the difference of two probabilities	35
1.2.4 Relative risks	39
1.2.5 Odds ratios	43
1.2.6 Matched pairs data	47
1.2.7 Larger contingency tables	51
1.3 Exercises	52
2 Analyzing a binary response, part 2: regression models	71
2.1 Linear regression models	71
2.2 Logistic regression models	72
2.2.1 Parameter estimation	75
2.2.2 Hypothesis tests for regression parameters	86
2.2.3 Odds ratios	93
2.2.4 Probability of success	98
2.2.5 Interactions and transformations for explanatory variables	109
2.2.6 Categorical explanatory variables	119
2.2.7 Convergence of parameter estimation	131
2.2.8 Monte Carlo simulation	138
2.3 Generalized linear models	143
2.4 Exercises	151
3 Analyzing a multcategory response	172
3.1 Multinomial probability distribution	172
3.2 $I \times J$ contingency tables and inference procedures	174
3.2.1 One multinomial distribution	174
3.2.2 I multinomial distributions	175
3.2.3 Test for independence	178
3.3 Nominal response regression models	184
3.3.1 Odds ratios	194

3.3.2	Contingency tables	201
3.4	Ordinal response regression models	209
3.4.1	Odds ratios	216
3.4.2	Contingency tables	220
3.4.3	Non-proportional odds model	223
3.5	Additional regression models	227
3.6	Exercises	227
4	Analyzing a count response	242
4.1	Poisson model for count data	242
4.1.1	Poisson distribution	242
4.1.2	Poisson likelihood and inference	246
4.2	Poisson regression models for count responses	251
4.2.1	Model for mean: log link	251
4.2.2	Parameter estimation and inference	252
4.2.3	Categorical explanatory variables	260
4.2.4	Poisson regression for contingency tables: loglinear models	267
4.2.5	Large loglinear models	273
4.2.6	Ordinal categorical variables	284
4.3	Poisson rate regression	296
4.4	Zero inflation	302
4.5	Exercises	312
5	Model selection and evaluation	331
5.1	Variable selection	331
5.1.1	Overview of variable selection	332
5.1.2	Model comparison criteria	333
5.1.3	All-subsets regression	335
5.1.4	Stepwise variable selection	340
5.1.5	Modern variable selection methods	346
5.1.6	Model averaging	350
5.2	Tools to assess model fit	354
5.2.1	Residuals	355
5.2.2	Goodness of fit	362
5.2.3	Influence	367
5.2.4	Diagnostics for multcategory response models	372
5.3	Overdispersion	373
5.3.1	Causes and implications	373
5.3.2	Detection	376
5.3.3	Solutions	378
5.4	Examples	390
5.4.1	Logistic regression: placekicking data set	390
5.4.2	Poisson regression: alcohol consumption data set	402
5.5	Inference after model selection	416
5.6	Exercises	423
6	Additional topics	435
6.1	Binary responses and testing error	435
6.1.1	Estimating the probability of success	435
6.1.2	Binary regression models	438
6.1.3	Other methods	441

6.2	Exact inference	442
6.2.1	Fisher's exact test for independence	442
6.2.2	Permutation test for independence	447
6.2.3	Exact logistic regression	452
6.2.4	Additional exact inference procedures	459
6.3	Categorical data analysis in complex survey designs	459
6.3.1	The survey sampling paradigm	459
6.3.2	Overview of analysis approaches	461
6.3.3	Weighted cell counts	464
6.3.4	Inference on population proportions	466
6.3.5	Contingency tables and loglinear models	468
6.3.6	Logistic regression	480
6.4	"Choose-all-that-apply" data	485
6.4.1	Item response table	486
6.4.2	Testing for marginal independence	486
6.4.3	Regression modeling	494
6.5	Mixed models and estimating equations for correlated data	502
6.5.1	Random effects	502
6.5.2	Mixed-effects models	504
6.5.3	Model fitting	508
6.5.4	Inference	515
6.5.5	Marginal modeling using generalized estimating equations	523
6.6	Bayesian methods for categorical data	528
6.6.1	Estimating a probability of success	529
6.6.2	Regression models	534
6.6.3	Alternative computational tools	545
6.7	Group testing	546
6.7.1	Identification	547
6.7.2	Estimation	556
6.7.3	Additional considerations and methods	563
6.8	Splines and additive models	564
6.8.1	Regression with splines	564
6.8.2	Generalized additive regression models	580
6.9	Exercises	581
A	An introduction to R	613
A.1	Basics	613
A.2	Functions	616
A.3	Help	617
A.4	Using functions on vectors	618
A.5	External packages	619
A.6	Programming environments	619
A.6.1	R editor	620
A.6.2	RStudio	620
A.6.3	Tinn-R	622
A.6.4	Other ways to use R	623
A.7	Regression example	626
A.7.1	Background	626
A.7.2	Data summary	629
A.7.3	Regression modeling	632
A.7.4	Additional items	637

B Likelihood methods	640
B.1 Introduction	640
B.1.1 Model and parameters	640
B.1.2 The role of maximum likelihood estimation	641
B.2 Likelihood	641
B.2.1 Definition	641
B.2.2 Examples	642
B.3 Maximum likelihood estimates	643
B.3.1 Mathematical maximization of the log-likelihood function	645
B.3.2 Computational maximization of the log-likelihood function	646
B.3.3 Large-sample properties of the MLE	647
B.3.4 Variance of the MLE	647
B.4 Functions of parameters	650
B.4.1 Invariance property of MLEs	650
B.4.2 Delta method for variances of functions	650
B.5 Inference with MLEs	651
B.5.1 Tests for parameters	651
B.5.2 Confidence intervals for parameters	655
B.5.3 Tests for models	656
Bibliography	659
Index	677