

Contents

Foreword	xix
Acknowledgments	xxi

CHAPTER 1 Introduction 1

1.1 Statistical models and why you need them	2
1.2 Why linear statistical models?	4
1.3 Why go beyond the linear model?	5
1.4 Random effects and why you need them	5
1.5 Why do you need both Bayesian and non-Bayesian statistical inference?	5
1.6 More reasons for why you should really understand maximum likelihood	6
1.7 The data simulation/model fitting dualism	6
1.8 The “experimental approach” to statistics	8
1.9 The first principle of modeling: start simple!	9
1.10 Overview of the ASM book and additional resources	9
1.11 How to use the ASM book for self-study and in courses	11
1.12 Summary and outlook	13

CHAPTER 2 Introduction to statistical inference 15

2.1 Probability as the basis for statistical inference	16
2.2 Random variables and their description by probability distributions	19
2.2.1 Discrete and continuous random variables	20
2.2.2 Description of random variables by probability distributions	20
2.2.3 Location and spread of a random variable, and what “modeling a parameter” means	26
2.2.4 Short summary on random variables and probability distributions	27
2.3 Statistical models and their usages	27
2.4 The likelihood function	30
2.5 Classical inference by maximum likelihood and its application to a single-parameter model	31
2.5.1 Introduction to maximum likelihood	32
2.5.2 Confidence intervals by inversion of the likelihood ratio test	38
2.5.3 Standard errors and confidence intervals based on approximate normality of the maximum likelihood estimates	40
2.5.4 Obtaining standard errors and confidence intervals using the bootstrap	41
2.5.5 A short summary on maximum likelihood estimation	45
2.6 Maximum likelihood estimation in a two-parameter model	45

2.7	Bayesian inference using posterior distributions.....	53
2.7.1	Probability as a general measure of uncertainty, or of imperfect knowledge	54
2.7.2	The posterior distribution as a fundamental target of inference.....	56
2.7.3	Prior distributions.....	58
2.8	Bayesian computation by Markov chain Monte Carlo (MCMC)	60
2.8.1	Monte Carlo integration.....	61
2.8.2	MCMC for the Swiss bee-eaters.....	63
2.8.3	A practical in MCMC for the bee-eaters with function <code>demoMCMC()</code> (*)	71
2.9	So should you now be a Bayesian or a frequentist?	73
2.10	Summary and outlook	74
	Further reading	76

CHAPTER 3	Linear regression models and their extensions to generalized linear, hierarchical, and integrated models	77
3.1	Introduction	77
3.2	Statistical distributions for the random variables in our model.....	78
3.2.1	Normal distribution.....	79
3.2.2	Uniform distribution	80
3.2.3	Binomial distribution: the “coin-flip distribution”	81
3.2.4	Poisson distribution.....	83
3.2.5	Summary remarks on statistical distributions	84
3.3	Link functions to model parameters on a transformed scale	85
3.4	Linear modeling of covariate effects.....	86
3.4.1	The “model of the mean”	88
3.4.2	Comparison of two groups.....	89
3.4.3	Simple linear regression: modeling the effect of one continuous covariate	92
3.4.4	Modeling the effects of one categorical covariate, or factor.....	94
3.4.5	Modeling the effects of two factors	96
3.4.6	Modeling the effects of a factor and a continuous covariate.....	100
3.5	Brief overview of linear, generalized linear, (generalized) linear mixed, hierarchical, and integrated models	106
3.6	Summary and outlook	107

CHAPTER 4	Introduction to general-purpose model fitting engines and the “model of the mean”	109
4.1	Introduction	109
4.2	Data generation	110
4.3	Analysis using canned functions in R	113
4.4	JAGS.....	113
4.4.1	Introduction to JAGS	113

4.4.2	Fit the model with JAGS	115
4.5	NIMBLE.....	125
4.5.1	Introduction to NIMBLE	125
4.5.2	Fit the model with NIMBLE	127
4.6	Stan.....	131
4.6.1	Introduction to Stan	131
4.6.2	Fit the model with Stan	132
4.7	Maximum likelihood in R.....	136
4.7.1	Introduction to maximum likelihood in R.....	136
4.7.2	Fit the model using maximum likelihood in R	137
4.8	Maximum likelihood using Template Model Builder (TMB)	141
4.8.1	Introduction to Template Model Builder	141
4.8.2	Fit the model using TMB	142
4.9	Comparison of engines and concluding remarks	146

CHAPTER 5 Normal linear regression..... 147

5.1	Introduction	147
5.2	Data generation	149
5.3	Analysis with canned functions in R.....	150
5.3.1	Fitting the model.....	150
5.3.2	Goodness-of-fit assessment using traditional and quantile residuals	151
5.3.3	Computing predictions.....	153
5.4	Bayesian analysis with JAGS	154
5.4.1	Fitting the model.....	155
5.4.2	Goodness-of-fit assessment in Bayesian analyses.....	157
5.4.3	Computing predictions.....	160
5.4.4	Interpretation of confidence interval versus credible intervals.....	162
5.4.5	And what about prediction intervals?.....	163
5.5	Bayesian analysis with NIMBLE	164
5.6	Bayesian analysis with Stan.....	164
5.7	Do-it-yourself maximum likelihood estimates	166
5.8	Likelihood analysis with TMB	167
5.9	Comparison of the parameter estimates	169
5.10	Summary and outlook	170

CHAPTER 6 Comparing two groups in a normal model..... 171

6.1	Introduction	171
6.2	Comparing two groups with equal variances	172
6.2.1	Data generation	173
6.2.2	Likelihood analysis with canned functions in R.....	174
6.2.3	Bayesian analysis with JAGS	175

6.2.4	Bayesian analysis with NIMBLE	177
6.2.5	Bayesian analysis with Stan	177
6.2.6	Do-it-yourself maximum likelihood estimates	180
6.2.7	Likelihood analysis with TMB	181
6.2.8	Comparison of the parameter estimates	182
6.3	Comparing two groups with unequal variances	182
6.3.1	Data generation	183
6.3.2	Frequentist analysis with canned functions in R	184
6.3.3	Bayesian analysis with JAGS	184
6.3.4	Bayesian analysis with NIMBLE	186
6.3.5	Bayesian analysis with Stan	186
6.3.6	Do-it-yourself maximum likelihood estimates	187
6.3.7	Likelihood analysis with TMB	188
6.3.8	Comparison of the parameter estimates	190
6.4	Summary and a comment on the modeling of variances	190

CHAPTER 7	Models with a single categorical covariate with more than two levels	191
7.1	Introduction: fixed and random effects	191
7.2	Fixed-effects models	195
7.2.1	Data generation	195
7.2.2	Likelihood analysis with canned functions in R	196
7.2.3	Bayesian analysis with JAGS	196
7.2.4	Bayesian analysis with NIMBLE	198
7.2.5	Bayesian analysis with Stan	199
7.2.6	Do-it-yourself maximum likelihood estimates	200
7.2.7	Likelihood analysis with TMB	201
7.2.8	Comparison of the parameter estimates	202
7.3	Random-effects models	202
7.3.1	Data generation	202
7.3.2	Likelihood analysis with canned functions in R	204
7.3.3	Bayesian analysis with JAGS ... and a quick fixed-random comparison	206
7.3.4	Bayesian analysis with NIMBLE	210
7.3.5	Bayesian analysis with Stan	210
7.3.6	Do-it-yourself maximum likelihood estimates	212
7.3.7	Likelihood analysis with TMB	214
7.3.8	Comparison of the parameter estimates	216
7.4	Summary and outlook	217

CHAPTER 8	Comparisons along two classifications in a model with two factors	219
8.1	Introduction: main and interaction effects	219
8.2	Data generation	221
8.3	Likelihood analysis with canned functions in R	223
8.4	An aside: using simulation to assess bias and precision of an estimator ... and to understand what a standard error is	225
8.5	Bayesian analysis with JAGS	228
8.5.1	Main-effects model	228
8.5.2	Interaction-effects model	230
8.5.3	Forming predictions	232
8.6	Bayesian analysis with NIMBLE	233
8.6.1	Main-effects model	233
8.6.2	Interaction-effects model	233
8.7	Bayesian analysis with Stan	233
8.7.1	Main-effects model	233
8.7.2	Interaction-effects model	235
8.8	Do-it-yourself maximum likelihood estimates	236
8.8.1	Main-effects model	236
8.8.2	Interaction-effects model	237
8.9	Likelihood analysis with TMB	238
8.9.1	Main-effects model	238
8.9.2	Interaction-effects model	239
8.10	Comparison of the parameter estimates	240
8.11	Summary and outlook	240
CHAPTER 9	General linear model for a normal response with continuous and categorical explanatory variables	243
9.1	Introduction	243
9.2	Data generation	245
9.3	Likelihood analysis with canned functions in R	246
9.4	Bayesian analysis with JAGS	248
9.5	Bayesian analysis with NIMBLE	250
9.6	Bayesian analysis with Stan	250
9.7	Do-it-yourself maximum likelihood estimates	252
9.8	Likelihood analysis with TMB	253
9.9	Comparison of the parameter estimates	255
9.10	Summary and outlook	255

CHAPTER 10 Linear mixed-effects model	257
10.1 Introduction	258
10.2 Data generation	260
10.3 Analysis under a random-intercepts model	262
10.3.1 Likelihood analysis using canned functions in R	262
10.3.2 Bayesian analysis with JAGS	263
10.3.3 Bayesian analysis with NIMBLE	266
10.3.4 Bayesian analysis with Stan	266
10.3.5 Do-it-yourself maximum likelihood estimates	267
10.3.6 Likelihood analysis with TMB	269
10.3.7 Comparison of the parameter estimates	271
10.4 Analysis under a random-coefficients model without correlation between intercept and slope	271
10.4.1 Likelihood analysis using canned functions in R	271
10.4.2 Bayesian analysis with JAGS	273
10.4.3 Bayesian analysis with NIMBLE	274
10.4.4 Bayesian analysis with Stan	275
10.4.5 Do-it-yourself maximum likelihood estimates	276
10.4.6 Likelihood analysis with TMB	278
10.4.7 Comparison of the parameter estimates	280
10.5 The random-coefficients model with correlation between intercept and slope	280
10.5.1 Introduction	280
10.5.2 Data generation	281
10.5.3 Likelihood analysis using canned functions in R	282
10.5.4 Bayesian analysis with JAGS	283
10.5.5 Bayesian analysis with NIMBLE	285
10.5.6 Bayesian analysis with Stan	285
10.5.7 Do-it-yourself maximum likelihood estimates ¹	287
10.5.8 Likelihood analysis with TMB	290
10.5.9 Comparison of the parameter estimates	292
10.6 Summary and outlook	293
CHAPTER 11 Introduction to the generalized linear model (GLM): comparing two groups in a Poisson regression	295
11.1 Introduction	295
11.2 An important but often forgotten issue with count data	298
11.3 How to deal with missing values in our data	298
11.4 Data generation	299
11.5 Likelihood analysis with canned functions in R	300

11.6	Bayesian analysis with JAGS	301
11.6.1	Assessment of model adequacy	303
11.6.2	Inference under the model	304
11.7	Bayesian analysis with NIMBLE	306
11.8	Bayesian analysis with Stan.....	306
11.9	Do-it-yourself maximum likelihood estimates	308
11.10	Likelihood analysis with TMB	309
11.11	Comparison of the parameter estimates	310
11.12	Summary and outlook	310

CHAPTER 12	Overdispersion, zero inflation, and offsets in a Poisson GLM.....	313
12.1	Introduction	314
12.2	Overdispersion.....	314
12.2.1	Data generation	315
12.2.2	Likelihood analysis with canned functions in R.....	316
12.2.3	Bayesian analysis with JAGS	316
12.2.4	Bayesian analysis with NIMBLE	318
12.2.5	Bayesian analysis with Stan	318
12.2.6	Do-it-yourself maximum likelihood estimates.....	319
12.2.7	Likelihood analysis with TMB	320
12.2.8	Comparison of the parameter estimates	321
12.3	Zero inflation.....	321
12.3.1	Data generation	323
12.3.2	Likelihood analysis with canned functions in R.....	323
12.3.3	Bayesian analysis with JAGS	324
12.3.4	Bayesian analysis with NIMBLE	325
12.3.5	Bayesian analysis with Stan	325
12.3.6	Do-it-yourself maximum likelihood estimates.....	326
12.3.7	Likelihood analysis with TMB	327
12.3.8	Comparison of the parameter estimates	329
12.4	Offsets.....	329
12.4.1	Data generation	330
12.4.2	Likelihood analysis with canned functions in R.....	330
12.4.3	Bayesian analysis with JAGS	331
12.4.4	Bayesian analysis with NIMBLE	332
12.4.5	Bayesian analysis with Stan	332
12.4.6	Do-it-yourself maximum likelihood estimates.....	333
12.4.7	Likelihood analysis with TMB	333
12.4.8	Comparison of the parameter estimates	334
12.5	Summary and outlook	335

CHAPTER 13 Poisson GLM with continuous and categorical explanatory variables.....	337
13.1 Introduction	337
13.2 Data generation	339
13.3 Likelihood analysis with canned functions in R	340
13.4 Bayesian analysis with JAGS	341
13.4.1 Fitting the model.....	341
13.4.2 Forming predictions	343
13.5 Bayesian analysis with NIMBLE	345
13.6 Bayesian analysis with Stan.....	345
13.7 Do-it-yourself maximum likelihood estimates	346
13.8 Likelihood analysis with TMB	347
13.9 Comparison of the parameter estimates	348
13.10 Summary.....	348
 CHAPTER 14 Poisson generalized linear mixed model, or Poisson GLMM.....	 349
14.1 Introduction	349
14.2 Data generation	351
14.3 Likelihood analysis with canned functions in R	353
14.4 Bayesian analysis with JAGS	354
14.5 Bayesian analysis with NIMBLE	356
14.6 Bayesian analysis with Stan.....	356
14.7 Do-it-yourself maximum likelihood estimates	357
14.8 Likelihood analysis with TMB	359
14.9 Comparison of the parameter estimates	361
14.10 Summary.....	361
 CHAPTER 15 Comparing two groups in a logistic regression model	 363
15.1 Introduction	363
15.2 Data generation	365
15.3 Likelihood analysis with canned functions in R	366
15.4 Bayesian analysis with JAGS	367
15.5 Bayesian analysis with NIMBLE	369
15.6 Bayesian analysis with Stan.....	369
15.7 Do-it-yourself maximum likelihood estimates	370
15.8 Likelihood analysis with TMB	371
15.9 Comparison of the parameter estimates	373
15.10 Summary.....	373

CHAPTER 16 Binomial GLM with continuous and categorical explanatory variables.....	375
16.1 Introduction	375
16.2 Data generation	377
16.3 Likelihood analysis with canned functions in R	379
16.4 Bayesian analysis with JAGS	379
16.5 Bayesian analysis with NIMBLE	382
16.6 Bayesian analysis with Stan.....	382
16.7 Do-it-yourself maximum likelihood estimates	384
16.8 Likelihood analysis with TMB	385
16.9 Comparison of the estimates.....	386
16.10 Summary.....	386
CHAPTER 17 Binomial generalized linear mixed model.....	387
17.1 Introduction	387
17.2 Data generation	388
17.3 Likelihood analysis with canned functions in R	391
17.4 Bayesian analysis with JAGS	392
17.5 Bayesian analysis with NIMBLE	394
17.6 Bayesian analysis with Stan.....	394
17.7 Do-it-yourself maximum likelihood estimates	396
17.8 Likelihood analysis with TMB	397
17.9 Comparison of the parameter estimates	399
17.10 Summary.....	399
CHAPTER 18 Model building, model checking, and model selection	401
18.1 Introduction	401
18.2 Why do we build a statistical model?	402
18.3 Ticklish rattlesnakes.....	404
18.4 How do we build a statistical model?	406
18.4.1 Model expansion and iterating on a single model	407
18.4.2 Causal, path, and structural equation models.....	408
18.4.3 How to build a predictive model.....	408
18.5 Model checking and goodness-of-fit testing	409
18.5.1 Traditional residual diagnostics.....	410
18.5.2 Bootstrapped and posterior predictive distributions.....	411
18.5.3 Cross-validation for goodness-of-fit assessments	419
18.5.4 Identifying and correcting for overdispersion.....	423
18.5.5 Measures of the absolute fit of a model.....	425
18.5.6 Parameter identifiability and robustness to assumption violations.....	426

18.6	Model selection	429
18.6.1	When to do model selection . . . and when not	429
18.6.2	Measuring the predictive performance of a model	431
18.6.3	Fully external model validation using “out-of-sample” data.....	432
18.6.4	Approximating external validation for “in-sample” data with cross-validation	434
18.6.5	Approximating external validation for “in-sample” data with information criteria: AIC, DIC, WAIC, and LOO-IC	435
18.7	Model averaging.....	440
18.8	Regularization: penalization, shrinkage, ridge, and lasso	442
18.9	Summary and outlook	444

CHAPTER 19 Occupancy models **447**

19.1	Introduction	447
19.2	Data generation	452
19.3	Likelihood analysis with canned functions in the R package unmarked	456
19.4	Bayesian analysis with JAGS	458
19.5	Bayesian analysis with NIMBLE	462
19.6	Bayesian analysis with Stan.....	462
19.7	Bayesian analysis with canned functions in the R package ubms	465
19.8	Do-it-yourself maximum likelihood estimates	466
19.9	Likelihood analysis with TMB	467
19.10	Comparison of the parameter estimates	469
19.11	Summary and outlook	469

CHAPTER 20 Integrated models **473**

20.1	Introduction	473
20.2	Data generation: simulating three abundance data sets with different observation/aggregation models.....	479
20.3	Fitting models to the three individual data sets first.....	481
20.3.1	Fitting a standard Poisson GLM in R and JAGS to data set 1.....	481
20.3.2	Fitting the zero-truncated Poisson GLM in R and JAGS to data set 2	483
20.3.3	Fitting the cloglog Bernoulli GLM in R and JAGS to data set 3	485
20.4	Fitting the integrated model to all three data sets simultaneously	487
20.4.1	Fitting the IM with a DIY likelihood function	487
20.4.2	Fitting the IM with JAGS	488
20.4.3	Fitting the IM with NIMBLE	489
20.4.4	Fitting the IM with Stan	489
20.4.5	Fitting the IM with TMB.....	491

20.4.6	Comparison of the parameter estimates for the IM	492
20.5	What do we gain by analyzing the joint likelihood in our analysis?	492
20.6	Summary and outlook	494

CHAPTER 21 Conclusion..... 497

21.1	Looking back.....	497
21.2	Looking ahead	499
21.2.1	Study design.....	500
21.2.2	Do-it-yourself Markov chain Monte Carlo	501
21.2.3	Machine learning	501
21.2.4	Hierarchical models	502

Bibliography	505
--------------------	-----

Index	519
-------------	-----

Student resources and helpful ancillaries have been prepared to aid learning and teaching, including all code included in the book available for download. Please visit the companion site for more details: <https://www.elsevier.com/books-and-journals/book-companion/9780443137150>