

# Contents

## Preface

## 1 Linear Models

### 1.1 A simple linear model

#### Simple least squares estimation

#### 1.1.1 Sampling properties of $\hat{\beta}$

#### 1.1.2 So how old is the universe?

#### 1.1.3 Adding a distributional assumption

#### Testing hypotheses about $\beta$

#### Confidence intervals

### 1.2 Linear models in general

### 1.3 The theory of linear models

#### 1.3.1 Least squares estimation of $\beta$

#### 1.3.2 The distribution of $\hat{\beta}$

#### 1.3.3 $(\hat{\beta}_i - \beta_i) / \hat{\sigma} \hat{\beta}_i \sim t_{n-p}$

#### 1.3.4 F-ratio results

#### 1.3.5 The influence matrix

#### 1.3.6 The residuals, $\hat{\epsilon}$ , and fitted values, $\hat{\mu}$

#### 1.3.7 Results in terms of $\mathbf{X}$

#### 1.3.8 The Gauss Markov Theorem: What's special about least squares?

### 1.4 The geometry of linear modelling

#### 1.4.1 Least squares

#### 1.4.2 Fitting by orthogonal decompositions

1.4.3	Comparison of nested models	21
1.5	Practical linear modelling	22
1.5.1	Model fitting and model checking	23
1.5.2	Model summary	28
1.5.3	Model selection	30
1.5.4	Another model selection example	31
A follow-up		35
1.5.5	Confidence intervals	36
1.5.6	Prediction	36
1.6	Practical modelling with factors	37
1.6.1	Identifiability	38
1.6.2	Multiple factors	39
1.6.3	'Interactions' of factors	40
1.6.4	Using factor variables in R	41
1.7	General linear model specification in R	44
1.8	Further linear modelling theory	45
1.8.1	Constraints I: General linear constraints	46
1.8.2	Constraints II: 'Contrasts' and factor variables	46
1.8.3	Likelihood	48
1.8.4	Non-independent data with variable variance	49
1.8.5	AIC and Mallows' statistic	51
1.8.6	Non-linear least squares	53
1.8.7	Further reading	55
1.9	Exercises	55
<b>2</b>	<b>Generalized Linear Models</b>	<b>59</b>
2.1	The theory of GLMs	60
2.1.1	The exponential family of distributions	62
2.1.2	Fitting generalized linear models	63
2.1.3	The IRLS objective is a quadratic approximation to the log-likelihood	66

2.1.4	AIC for GLMs	68
2.1.5	Large sample distribution of $\hat{\beta}$	69
2.1.6	Comparing models by hypothesis testing	69
	Deviance	70
	Model comparison with unknown $\phi$	71
2.1.7	$\hat{\phi}$ and Pearson's statistic	71
2.1.8	Canonical link functions	72
2.1.9	Residuals	73
	Pearson residuals	73
	Deviance residuals	73
2.1.10	Quasi-likelihood	74
2.2	Geometry of GLMs	76
2.2.1	The geometry of IRLS	77
2.2.2	Geometry and IRLS convergence	78
2.3	GLMs with R	81
2.3.1	Binomial models and heart disease	81
2.3.2	A Poisson regression epidemic model	87
2.3.3	Log-linear models for categorical data	93
2.3.4	Sole eggs in the Bristol channel	97
2.4	Likelihood	102
2.4.1	Invariance	102
2.4.2	Properties of the expected log-likelihood	103
2.4.3	Consistency	106
2.4.4	Large sample distribution of $\hat{\theta}$	107
2.4.5	The generalized likelihood ratio test (GLRT)	108
2.4.6	Derivation of $2\lambda \sim \chi_r^2$ under $H_0$	109
2.4.7	AIC in general	111
2.4.8	Quasi-likelihood results	113
2.5	Exercises	115

<b>3</b>	<b>Introducing GAMs</b>	<b>121</b>
3.1	Introduction	121
3.2	Univariate smooth functions	122
3.2.1	Representing a smooth function: Regression splines	122
	A very simple example: A polynomial basis	122
	Another example: A cubic spline basis	124
	Using the cubic spline basis	126
3.2.2	Controlling the degree of smoothing with penalized regression splines	128
3.2.3	Choosing the smoothing parameter, $\lambda$ : Cross validation	130
3.3	Additive models	133
3.3.1	Penalized regression spline representation of an additive model	134
3.3.2	Fitting additive models by penalized least squares	135
3.4	Generalized additive models	137
3.5	Summary	139
3.6	Exercises	140
<b>4</b>	<b>Some GAM Theory</b>	<b>145</b>
4.1	Smoothing bases	146
4.1.1	Why splines?	146
	Natural cubic splines are smoothest interpolators	146
	Cubic smoothing splines	148
4.1.2	Cubic regression splines	149
4.1.3	A cyclic cubic regression spline	151
4.1.4	P-splines	152
4.1.5	Thin plate regression splines	154
	Thin plate splines	154
	Thin plate regression splines	157
	Properties of thin plate regression splines	158
	Knot-based approximation	160
4.1.6	Shrinkage smoothers	160

4.1.7	Choosing the basis dimension	161
4.1.8	Tensor product smooths	162
	Tensor product bases	162
	Tensor product penalties	165
4.2	Setting up GAMs as penalized GLMs	167
4.2.1	Variable coefficient models	168
4.3	Justifying P-IRLS	169
4.4	Degrees of freedom and residual variance estimation	170
4.4.1	Residual variance or scale parameter estimation	171
4.5	Smoothing parameter selection criteria	172
4.5.1	Known scale parameter: UBRE	172
4.5.2	Unknown scale parameter: Cross validation	173
	Problems with ordinary cross validation	174
4.5.3	Generalized cross validation	175
4.5.4	GCV/UBRE/AIC in the generalized case	177
	Approaches to GAM GCV/UBRE minimization	179
4.6	Numerical GCV/UBRE: Performance iteration	181
4.6.1	Minimizing the GCV or UBRE score	181
	Stable and efficient evaluation of the scores and derivatives	183
	The weighted constrained case	185
4.7	Numerical GCV/UBRE optimization by outer iteration	186
4.7.1	Differentiating the GCV/UBRE function	187
4.8	Distributional results	189
4.8.1	Bayesian model, and posterior distribution of the parameters, for an additive model	190
4.8.2	Structure of the prior	191
4.8.3	Posterior distribution for a GAM	192
4.8.4	Bayesian confidence intervals for non-linear functions of parameters	194
4.8.5	P-values	194
4.9	Confidence interval performance	196

4.9.1	Single smooths	196
4.9.2	GAMs and their components	200
4.9.3	Unconditional Bayesian confidence intervals	202
4.10	Further GAM theory	204
4.10.1	Comparing GAMs by hypothesis testing	204
4.10.2	ANOVA decompositions and nesting	206
4.10.3	The geometry of penalized regression	208
4.10.4	The “natural” parameterization of a penalized smoother	210
4.11	Other approaches to GAMs	212
4.11.1	Backfitting GAMs	213
4.11.2	Generalized smoothing splines	215
4.12	Exercises	217
<b>5</b>	<b>GAMs in Practice: mgcv</b>	<b>221</b>
5.1	Cherry trees again	221
5.1.1	Finer control of gam	223
5.1.2	Smooths of several variables	225
5.1.3	Parametric model terms	228
5.2	Brain imaging example	230
5.2.1	Preliminary modelling	232
5.2.2	Would an additive structure be better?	236
5.2.3	Isotropic or tensor product smooths?	237
5.2.4	Detecting symmetry (with by variables)	239
5.2.5	Comparing two surfaces	241
5.2.6	Prediction with <code>predict.gam</code>	243
	Prediction with <code>lpmatrix</code>	245
5.2.7	Variances of non-linear functions of the fitted model	246
5.3	Air pollution in Chicago example	247
5.4	Mackerel egg survey example	254
5.4.1	Model development	254
5.4.2	Model predictions	260

CONTENTS	xiii
5.5 Portuguese larks example	262
5.6 Other packages	265
5.6.1 Package <code>gam</code>	265
5.6.2 Package <code>gss</code>	267
5.7 Exercises	270
<b>6 Mixed Models and GAMMs</b>	<b>277</b>
6.1 Mixed models for balanced data	277
6.1.1 A motivating example	277
The wrong approach: A fixed effects linear model	278
The right approach: A mixed effects model	280
6.1.2 General principles	281
6.1.3 A single random factor	282
6.1.4 A model with two factors	286
6.1.5 Discussion	290
6.2 Linear mixed models in general	291
6.2.1 Estimation of linear mixed models	292
6.2.2 Directly maximizing a mixed model likelihood in R	293
6.2.3 Inference with linear mixed models	295
Fixed effects	295
Inference about the random effects	296
6.2.4 Predicting the random effects	297
6.2.5 REML	298
The explicit form of the REML criterion	299
6.2.6 A link with penalized regression	300
6.2.7 The EM algorithm	302
6.3 Linear mixed models in R	303
6.3.1 Tree growth: An example using <code>lme</code>	304
6.3.2 Several levels of nesting	309
6.4 Generalized linear mixed models	310
6.5 GLMMs with R	312

6.6	Generalized additive mixed models	316
6.6.1	Smooths as mixed model components	316
6.6.2	Inference with GAMMs	318
6.7	GAMMs with R	319
6.7.1	A GAMM for sole eggs	319
6.7.2	The temperature in Cairo	321
6.8	Exercises	325
<b>A</b>	<b>Some Matrix Algebra</b>	<b>331</b>
A.1	Basic computational efficiency	331
A.2	Covariance matrices	332
A.3	Differentiating a matrix inverse	332
A.4	Kronecker product	333
A.5	Orthogonal matrices and Householder matrices	333
A.6	QR decomposition	334
A.7	Choleski decomposition	334
A.8	Eigen-decomposition	335
A.9	Singular value decomposition	336
A.10	Pivoting	337
A.11	Lanczos iteration	337
<b>B</b>	<b>Solutions to Exercises</b>	<b>341</b>
B.1	Chapter 1	341
B.2	Chapter 2	346
B.3	Chapter 3	351
B.4	Chapter 4	353
B.5	Chapter 5	360
B.6	Chapter 6	369
	<b>Bibliography</b>	<b>379</b>
	<b>Index</b>	<b>385</b>