

Contents

Preface	v
Acknowledgements	vi
Datasets used in this book	vi
Cover art	vi
Contributors	xi
1 Review of multiple linear regression	1
1.1 Light levels and size of the human visual system	1
1.2 The variables	1
1.3 A protocol for the analysis	3
1.4 Data exploration	5
1.5 Multiple linear regression	9
1.5.1 Underlying statistical theory	9
1.5.2 Multiple linear regression	20
1.5.3 Fitting the model in R and estimated parameters	20
1.6 Finding the optimal model	21
1.7 Degrees of freedom	22
1.8 Model validation	23
1.9. Model interpretation	27
1.10 What happens if we use collinear covariates?	28
1.11 Should we have applied a mixed effects model?	29
1.12 What to do if things go wrong	30
1.13 What to present in a paper	31
2 Introduction to additive models using deep-sea fisheries data	33
2.1 Impact of deep-sea fisheries	33
2.2 First encounter with smoothers	34
2.2.1 Applying linear regression	34
2.2.2 Applying cubic polynomials	38
2.2.3 A simple GAM	39
2.2.4 Moving average and LOESS smoothers	40
2.3 Applying GAM in R using the mgcv package	44
2.4 Cross-validation	46
2.5 Model validation	50
2.5.1 Normality and homogeneity	50
2.5.2 Independence	51
2.6 Extending the GAM with more covariates	55
2.6.1 GAM with a smoother and a nominal covariate	55
2.6.2 GAM with an interaction term; first implementation	58
2.6.3 GAM with an interaction term; second implementation	61
2.6.4 GAM with an interaction term; third implementation	61
2.7 Transforming the density data	63
2.8 Allowing for heterogeneity	64
2.9 Transforming and allowing for heterogeneity	67
2.10 What to present in a paper	68

3 Technical aspects of GAM using pelagic bioluminescent organisms	69
3.1 Pelagic bioluminescent organism data	69
3.2 Linear regression	69
3.3 Polynomial regression model	72
3.4 Linear spline regression	73
3.5 Quadratic spline regression	78
3.6 Cubic regression splines	80
3.7 The number of knots	84
3.8 Penalized quadratic spline regression	84
3.9 Other smoothers	88
3.10 Cubic smoothing spline	89
3.11 Summary of smoother types	93
3.12 Degrees of freedom of a smoother*	94
3.13 Bias-variance trade-off	96
3.14 Confidence intervals	96
3.15 Using the function gam in mgcv	99
3.16 The danger of using GAM	102
3.17 Additive models with multiple smoothers	106
4 Introducing generalized additive models using deep-sea fishery data	111
4.1 From additive models to generalized additive models	111
4.2 Review of GLM	111
4.2.1 Distribution	112
4.2.2 Predictor function	114
4.2.3 Link function	114
4.3 Start with GLM or GAM?	114
4.4 Results of Poisson and negative binomial GLM	115
4.5 Using the offset in a GLM or GAM	118
4.6 Poisson GLM with offset	119
4.7 Negative binomial GLM with offset	120
4.8 Poisson and negative binomial GAM with offset	120
4.9 What to present in paper	126
5 Additive modelling applied on stable isotope ratios of ocean squid	127
5.1 Stable isotope ratios of squid	127
5.2 The variables	128
5.3 Data exploration	129
5.4 Brainstorming	131
5.5 Applying the multiple linear regression model	132
5.6 Applying an additive model	134
5.7 Testing linearity versus non-linearity	136
5.7.1 Programming a smoother manually**	137
5.7.2 Summary of the mathematics	141
5.8 Consequences of ignoring collinearity in the additive model	141
5.9 Discussion	142
5.10 What to present in a paper	143

6 Generalized Additive Models applied on northern gannets	145
6.1 Northern gannets in the North Sea	145
6.2 The variables	145
6.3 Brainstorming	146
6.4 Data exploration	150
6.5 Building up the complexity of the GAMs	152
6.6 Zero-inflated GAM	161
6.6.1 A zero-inflated model for the gannet data	162
6.6.2 ZIP GAM using gamlss	165
6.7 Discussion	167
6.8 What to present in a paper	168
7 Generalized Additive Models applied on parasites of Argentine hake	169
7.1 Parasites of Argentine hake in the Argentine Sea	169
7.2 The variables	169
7.3 Data exploration	171
7.4 Brainstorming	173
7.5 Applying binomial GAM	174
7.6 Discussion	183
7.7 What to present in a paper	184
References	185
Index	189

