
Table of Contents

1	Introduction	1
1.1	Examples of Applications	4
1.2	First Steps	11
1.2.1	Univariate Distributions of the Variables	11
1.2.2	Graphical Association Analysis	13
1.3	Notational Remarks	19
2	Regression Models	21
2.1	Introduction	21
2.2	Linear Regression Models	22
2.2.1	Simple Linear Regression Model	22
2.2.2	Multiple Linear Regression	26
2.3	Regression with Binary Response Variables: The Logit Model....	33
2.4	Mixed Models	38
2.5	Simple Nonparametric Regression	44
2.6	Additive Models	49
2.7	Generalized Additive Models	52
2.8	Geoadditive Regression	55
2.9	Beyond Mean Regression	61
2.9.1	Regression Models for Location, Scale, and Shape	62
2.9.2	Quantile Regression	66
2.10	Models in a Nutshell	68
2.10.1	Linear Models (LMs, Chaps. 3 and 4)	68
2.10.2	Logit Model (Chap. 5)	68
2.10.3	Poisson Regression (Chap. 5)	68
2.10.4	Generalized Linear Models (GLMs, Chaps. 5 and 6)	69
2.10.5	Linear Mixed Models (LMMs, Chap. 7)	69
2.10.6	Additive Models and Extensions (AMs, Chaps. 8 and 9)	70
2.10.7	Generalized Additive (Mixed) Models (GA(M)Ms, Chap. 9)	70
2.10.8	Structured Additive Regression (STAR, Chap. 9)	71
2.10.9	Quantile Regression (Chap. 10)	71

3	The Classical Linear Model	73
3.1	Model Definition	73
3.1.1	Model Parameters, Estimation, and Residuals	77
3.1.2	Discussion of Model Assumptions	78
3.1.3	Modeling the Effects of Covariates	86
3.2	Parameter Estimation	104
3.2.1	Estimation of Regression Coefficients	104
3.2.2	Estimation of the Error Variance	108
3.2.3	Properties of the Estimators	110
3.3	Hypothesis Testing and Confidence Intervals	125
3.3.1	Exact F-Test	128
3.3.2	Confidence Regions and Prediction Intervals	136
3.4	Model Choice and Variable Selection	139
3.4.1	Bias, Variance and Prediction Quality	142
3.4.2	Model Choice Criteria	146
3.4.3	Practical Use of Model Choice Criteria	150
3.4.4	Model Diagnosis	155
3.5	Bibliographic Notes and Proofs	168
3.5.1	Bibliographic Notes	168
3.5.2	Proofs	168
4	Extensions of the Classical Linear Model	177
4.1	The General Linear Model	177
4.1.1	Model Definition	177
4.1.2	Weighted Least Squares	178
4.1.3	Heteroscedastic Errors	182
4.1.4	Autocorrelated Errors	191
4.2	Regularization Techniques	201
4.2.1	Statistical Regularization	202
4.2.2	Ridge Regression	203
4.2.3	Least Absolute Shrinkage and Selection Operator	208
4.2.4	Geometric Properties of Regularized Estimates	211
4.2.5	Partial Regularization	216
4.3	Boosting Linear Regression Models	217
4.3.1	Basic Principles	217
4.3.2	Componentwise Boosting	218
4.3.3	Generic Componentwise Boosting	222
4.4	Bayesian Linear Models	225
4.4.1	Standard Conjugate Analysis	227
4.4.2	Regularization Priors	237
4.4.3	Classical Bayesian Model Choice (and Beyond)	243
4.4.4	Spike and Slab Priors	253
4.5	Bibliographic Notes and Proofs	257
4.5.1	Bibliographic Notes	257
4.5.2	Proofs	258

5	Generalized Linear Models	269
5.1	Binary Regression	270
5.1.1	Binary Regression Models	270
5.1.2	Maximum Likelihood Estimation	279
5.1.3	Testing Linear Hypotheses	285
5.1.4	Criteria for Model Fit and Model Choice	287
5.1.5	Estimation of the Overdispersion Parameter	292
5.2	Count Data Regression	293
5.2.1	Models for Count Data	293
5.2.2	Estimation and Testing: Likelihood Inference	295
5.2.3	Criteria for Model Fit and Model Choice	297
5.2.4	Estimation of the Overdispersion Parameter	297
5.3	Models for Nonnegative Continuous Response Variables	298
5.4	Generalized Linear Models	301
5.4.1	General Model Definition	301
5.4.2	Likelihood Inference	306
5.5	Quasi-likelihood Models	309
5.6	Bayesian Generalized Linear Models	311
5.6.1	Posterior Mode Estimation	313
5.6.2	Fully Bayesian Inference via MCMC Simulation Techniques	314
5.6.3	MCMC-Based Inference Using Data Augmentation	316
5.7	Boosting Generalized Linear Models	319
5.8	Bibliographic Notes and Proofs	320
5.8.1	Bibliographic Notes	320
5.8.2	Proofs	321
6	Categorical Regression Models	325
6.1	Introduction	325
6.2	Models for Unordered Categories	329
6.3	Ordinal Models	334
6.3.1	The Cumulative Model	334
6.3.2	The Sequential Model	337
6.4	Estimation and Testing: Likelihood Inference	343
6.5	Bibliographic Notes	347
7	Mixed Models	349
7.1	Linear Mixed Models for Longitudinal and Clustered Data	350
7.1.1	Random Intercept Models	350
7.1.2	Random Coefficient or Slope Models	357
7.1.3	General Model Definition and Matrix Notation	361
7.1.4	Conditional and Marginal Formulation	365
7.1.5	Stochastic Covariates	366
7.2	General Linear Mixed Models	368

7.3	Likelihood Inference in LMMs	371
7.3.1	Known Variance–Covariance Parameters	371
7.3.2	Unknown Variance–Covariance Parameters	372
7.3.3	Variability of Fixed and Random Effects Estimators	378
7.3.4	Testing Hypotheses	380
7.4	Bayesian Linear Mixed Models	383
7.4.1	Estimation for Known Covariance Structure	384
7.4.2	Estimation for Unknown Covariance Structure	385
7.5	Generalized Linear Mixed Models	389
7.5.1	GLMMs for Longitudinal and Clustered Data	389
7.5.2	Conditional and Marginal Models	392
7.5.3	GLMMs in General Form	394
7.6	Likelihood and Bayesian Inference in GLMMs	394
7.6.1	Penalized Likelihood and Empirical Bayes Estimation...	395
7.6.2	Fully Bayesian Inference Using MCMC	397
7.7	Practical Application of Mixed Models	401
7.7.1	General Guidelines and Recommendations	401
7.7.2	Case Study on Sales of Orange Juice	403
7.8	Bibliographic Notes and Proofs	409
7.8.1	Bibliographic Notes	409
7.8.2	Proofs	410
8	Nonparametric Regression	413
8.1	Univariate Smoothing	415
8.1.1	Polynomial Splines	415
8.1.2	Penalized Splines (P-Splines)	431
8.1.3	General Penalization Approaches	446
8.1.4	Smoothing Splines	448
8.1.5	Random Walks	452
8.1.6	Kriging	453
8.1.7	Local Smoothing Procedures	460
8.1.8	General Scatter Plot Smoothing	468
8.1.9	Choosing the Smoothing Parameter	478
8.1.10	Adaptive Smoothing Approaches	490
8.2	Bivariate and Spatial Smoothing	500
8.2.1	Tensor Product P-Splines	503
8.2.2	Radial Basis Functions and Thin Plate Splines	512
8.2.3	Kriging: Spatial Smoothing with Continuous Location Variables	515
8.2.4	Markov Random Fields	521
8.2.5	Summary of Roughness Penalty Approaches	527
8.2.6	Local and Adaptive Smoothing	529
8.3	Higher-Dimensional Smoothing	530
8.4	Bibliographic Notes	531

9	Structured Additive Regression	535
9.1	Additive Models	536
9.2	Geoadditive Regression	540
9.3	Models with Interactions	543
9.3.1	Models with Varying Coefficient Terms	544
9.3.2	Interactions Between Two Continuous Covariates	547
9.4	Models with Random Effects	549
9.5	Structured Additive Regression	553
9.6	Inference	561
9.6.1	Penalized Least Squares or Likelihood Estimation	561
9.6.2	Inference Based on Mixed Model Representation	566
9.6.3	Bayesian Inference Based on MCMC	568
9.7	Boosting STAR Models	573
9.8	Case Study: Malnutrition in Zambia	576
9.8.1	General Guidelines	576
9.8.2	Descriptive Analysis	580
9.8.3	Modeling Variants	583
9.8.4	Estimation Results and Model Evaluation	584
9.8.5	Automatic Function Selection	589
9.9	Bibliographic Notes	594
10	Quantile Regression	597
10.1	Quantiles	599
10.2	Linear Quantile Regression	601
10.2.1	Classical Quantile Regression	601
10.2.2	Bayesian Quantile Regression	609
10.3	Additive Quantile Regression	612
10.4	Bibliographic Notes and Proofs	616
10.4.1	Bibliographic Notes	616
10.4.2	Proofs	618
A	Matrix Algebra	621
A.1	Definition and Elementary Matrix Operations	621
A.2	Rank of a Matrix	626
A.3	Block Matrices and the Matrix Inversion Lemma	628
A.4	Determinant and Trace of a Matrix	629
A.5	Generalized Inverse	631
A.6	Eigenvalues and Eigenvectors	631
A.7	Quadratic Forms	633
A.8	Differentiation of Matrix Functions	635
B	Probability Calculus and Statistical Inference	639
B.1	Some Univariate Distributions	639
B.2	Random Vectors	645

B.3	Multivariate Normal Distribution	648
B.3.1	Definition and Properties	648
B.3.2	The Singular Multivariate Normal Distribution	650
B.3.3	Distributions of Quadratic Forms	651
B.3.4	Multivariate t-Distribution	651
B.3.5	Normal-Inverse Gamma Distribution	652
B.4	Likelihood Inference	653
B.4.1	Maximum Likelihood Estimation	653
B.4.2	Numerical Computation of the MLE	660
B.4.3	Asymptotic Properties of the MLE	662
B.4.4	Likelihood-Based Tests of Linear Hypotheses	662
B.4.5	Model Choice	664
B.5	Bayesian Inference	665
B.5.1	Basic Concepts of Bayesian Inference	665
B.5.2	Point and Interval Estimation	669
B.5.3	MCMC Methods	670
B.5.4	Model Selection	676
B.5.5	Model Averaging	679
Bibliography		681
Index		691