

# Springer Series in Statistics

For further volumes:

<http://www.springer.com/series/692>



Jon Wakefield

# Bayesian and Frequentist Regression Methods

 Springer

Jon Wakefield  
Departments of Statistics and Biostatistics  
University of Washington  
Seattle, Washington  
USA

ISSN 0172-7397

ISBN 978-1-4419-0924-4

ISBN 978-1-4419-0925-1 (eBook)

DOI 10.1007/978-1-4419-0925-1

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012952935

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*In the order of my meeting them, this book is  
dedicated to:*

*Norma Maureen Wakefield*

*Eric Louis Wakefield*

*Samantha Louise Wakefield*

*Felicity Zoe Moodie*

*Eleanor Anna Wakefield*

*Eric Stephen Wakefield*



# Preface

The past 25 years have seen great advances in both Bayesian and frequentist methods for data analysis. The most significant advance for the Bayesian approach has been the development of Markov chain Monte Carlo methods for estimating expectations with respect to the posterior, hence allowing flexible inference and routine implementation for a wide range of models. In particular, this development has led to the more widespread use of hierarchical models for dependent data. With respect to frequentist methods, estimating functions have emerged as a unifying approach for determining the properties of estimators. Generalized estimating equations provide a particularly important example of this methodology that allows inference for dependent data.

The aim of this book is to provide a modern description of Bayesian and frequentist methods of regression analysis and to illustrate the use of these methods on real data. Many books describe one or the other of the Bayesian or frequentist approaches to regression modeling in different contexts, and many mathematical statistics texts describe the theory behind Bayesian and frequentist approaches without providing a detailed description of specific methods. References to such texts are given at the end of Chaps. 2 and 3. Bayesian and frequentist methods are not viewed here as competitive, but rather as complementary techniques, and in this respect this book has some uniqueness.

In embarking on the writing of this book, I have been influenced by many current and former colleagues. My early training was in the Mathematics Department at the University of Nottingham and my first permanent academic teaching position was in the Mathematics Department at Imperial College of Science, Technology and Medicine in London. During this period I was introduced to the Bayesian paradigm and was greatly influenced by Adrian Smith, both as a lecturer and as a Ph.D. adviser. I have also benefited, and continue to benefit, from numerous conversations with Dave Stephens who I have known for over 25 years. Following my move to the University of Washington in Seattle I was exposed to a very modern view of frequentist methods in the Department of Biostatistics. In particular, Scott Emerson, Patrick Heagerty and Thomas Lumley have provided constant stimulation. These interactions, among many others, have influenced the way I now think about

statistics, and it is this exposure which I hope has allowed me to write a balanced account of Bayesian and frequentist methods. There is some theory in this book and some data analysis, but the focus is on material that lies between these endeavors and concerns methods. At the University of Washington there is an advanced three-course regression methods sequence and this book arose out of my teaching of the three courses in the sequence.

If modern computers had been available a 100 years ago, the discipline of statistics would have developed in a dramatically different fashion to the way in which it actually evolved. In particular, there would probably be less dependence on linear and generalized linear models, which are mathematically and computationally convenient. While these model classes are still useful and do possess a number of convenient mathematical and computational properties, I believe they should be viewed as just two choices within a far wider range of models that are now available. The approach to modeling that is encouraged in this book is to first specify the model suggested by the background science and to then proceed to examining the mathematical and computational aspects of the model.

As a preparation for this book, the reader is assumed to have a grasp of calculus and linear algebra and have taken first courses in probability and statistical theory. The content of this book is as follows. An introductory chapter describes a number of motivating examples and discusses general issues that need consideration before a regression analysis is carried out. This book is then broken into five parts: I, Inferential Approaches; II, Independent Data; III, Dependent Data; IV, Nonparametric Modeling; V, Appendices. The first two chapters of Part I provide descriptions of the frequentist and Bayesian approaches to inference, with a particular emphasis on the rationale of each approach and a delineation of situations in which one or the other approach is preferable. The third chapter in Part I discusses model selection and hypothesis testing. Part II considers independent data and contains three chapters on the linear model, general regression models (including generalized linear models), and binary data models. The two chapters of Part III consider dependent data with linear models and general regression models. Mixed models and generalized estimating equations are the approaches to inference that are emphasized. Part IV contains three chapters on nonparametric modeling with an emphasis on spline and kernel methods. The examples and simulation studies of this book were almost exclusively carried out within the freely available R programming environment. The code for the examples and figures may be found at:

<http://faculty.washington.edu/jonno/regression-methods.html>

along with the inevitable errata and links to datasets. Exercises are included at the end of all chapters but the first. Many of these exercises concern analyses of real data. In my own experience, a full understanding of methods requires their implementation and application to data.

In my own teaching I have based three one-quarter courses on the following. *Regression Methods for Independent Data* is based on Part II, dipping into topics in Part I as needed and using motivating examples from Chap. 1. *Regression Methods*

for *Dependent Data* centers on Part II, again using examples from Chap. 1, and building on the independent data material. Finally, *Nonparametric Regression and Classification* is based on the material in Part IV. The latter course is stand-alone in the sense of not requiring the independent and dependent data courses though extra material on a number of topics, including linear and generalized linear models and mixed models, will need to be included if not previously encountered.

In the 2003–2004 academic year I was the Genentech Professor and received funding specifically to work on this book. The staff at Springer have been very helpful at all stages. John Kimmel was the editor during most of the writing of this book and I am appreciative of his gentle prodding and advice. About 18 months from the completion of this book, Marc Strauss stepped in and has also been very supportive. Many of my colleagues have given comments on various chapters, but I would like to specifically thank Lurdes Inoue, Katie Kerr, Erica Moodie, Zoe Moodie, Ken Rice, Dave Stephens, Jon Wellner, Daniela Witten, and Simon Wood for feedback on different parts of this book. Finally, lest we forget, I would like to thank all of those students who suffered through initial presentations of this material—I hope your sacrifices were not in vain. . .

Seattle, WA  
June 2012

Jon Wakefield



# Contents

<b>1</b>	<b>Introduction and Motivating Examples</b> .....	1
1.1	Introduction .....	1
1.2	Model Formulation .....	1
1.3	Motivating Examples .....	5
1.3.1	Prostate Cancer .....	5
1.3.2	Outcome After Head Injury .....	9
1.3.3	Lung Cancer and Radon .....	10
1.3.4	Pharmacokinetic Data .....	12
1.3.5	Dental Growth .....	16
1.3.6	Spinal Bone Mineral Density .....	18
1.4	Nature of Randomness .....	20
1.5	Bayesian and Frequentist Inference .....	22
1.6	The Executive Summary .....	23
1.7	Bibliographic Notes .....	24

## Part I Inferential Approaches

<b>2</b>	<b>Frequentist Inference</b> .....	27
2.1	Introduction .....	27
2.2	Frequentist Criteria .....	29
2.3	Estimating Functions .....	32
2.4	Likelihood .....	36
2.4.1	Maximum Likelihood Estimation .....	36
2.4.2	Variants on Likelihood .....	44
2.4.3	Model Misspecification .....	46
2.5	Quasi-likelihood .....	49
2.5.1	Maximum Quasi-likelihood Estimation .....	49
2.5.2	A More Complex Mean–Variance Model .....	53
2.6	Sandwich Estimation .....	56
2.7	Bootstrap Methods .....	63
2.7.1	The Bootstrap for a Univariate Parameter .....	64

2.7.2	The Bootstrap for Regression .....	66
2.7.3	Sandwich Estimation and the Bootstrap .....	66
2.8	Choice of Estimating Function .....	70
2.9	Hypothesis Testing .....	72
2.9.1	Motivation .....	72
2.9.2	Preliminaries .....	73
2.9.3	Score Tests .....	74
2.9.4	Wald Tests .....	75
2.9.5	Likelihood Ratio Tests .....	75
2.9.6	Quasi-likelihood .....	76
2.9.7	Comparison of Test Statistics .....	77
2.10	Concluding Remarks .....	79
2.11	Bibliographic Notes .....	80
2.12	Exercises .....	80
<b>3</b>	<b>Bayesian Inference</b> .....	<b>85</b>
3.1	Introduction .....	85
3.2	The Posterior Distribution and Its Summarization .....	86
3.3	Asymptotic Properties of Bayesian Estimators .....	89
3.4	Prior Choice .....	90
3.4.1	Baseline Priors .....	90
3.4.2	Substantive Priors .....	93
3.4.3	Priors on Meaningful Scales .....	95
3.4.4	Frequentist Considerations .....	96
3.5	Model Misspecification .....	99
3.6	Bayesian Model Averaging .....	100
3.7	Implementation .....	102
3.7.1	Conjugacy .....	102
3.7.2	Laplace Approximation .....	106
3.7.3	Quadrature .....	107
3.7.4	Integrated Nested Laplace Approximations .....	109
3.7.5	Importance Sampling Monte Carlo .....	110
3.7.6	Direct Sampling Using Conjugacy .....	112
3.7.7	Direct Sampling Using the Rejection Algorithm .....	114
3.8	Markov Chain Monte Carlo .....	121
3.8.1	Markov Chains for Exploring Posterior Distributions ...	121
3.8.2	The Metropolis–Hastings Algorithm .....	122
3.8.3	The Metropolis Algorithm .....	123
3.8.4	The Gibbs Sampler .....	123
3.8.5	Combining Markov Kernels: Hybrid Schemes .....	125
3.8.6	Implementation Details .....	125
3.8.7	Implementation Summary .....	133
3.9	Exchangeability .....	134
3.10	Hypothesis Testing with Bayes Factors .....	137
3.11	Bayesian Inference Based on a Sampling Distribution .....	140
3.12	Concluding Remarks .....	143

- 3.13 Bibliographic Notes..... 145
- 3.14 Exercises ..... 145
- 4 Hypothesis Testing and Variable Selection ..... 153**
  - 4.1 Introduction ..... 153
  - 4.2 Frequentist Hypothesis Testing ..... 153
    - 4.2.1 Fisherian Approach ..... 154
    - 4.2.2 Neyman–Pearson Approach ..... 154
    - 4.2.3 Critique of the Fisherian Approach..... 154
    - 4.2.4 Critique of the Neyman–Pearson Approach ..... 155
  - 4.3 Bayesian Hypothesis Testing with Bayes Factors ..... 156
    - 4.3.1 Overview of Approaches..... 156
    - 4.3.2 Critique of the Bayes Factor Approach ..... 158
    - 4.3.3 A Bayesian View of Frequentist Hypothesis Testing .... 159
  - 4.4 The Jeffreys–Lindley Paradox..... 161
  - 4.5 Testing Multiple Hypotheses: General Considerations ..... 164
  - 4.6 Testing Multiple Hypotheses: Fixed Number of Tests ..... 165
    - 4.6.1 Frequentist Analysis ..... 166
    - 4.6.2 Bayesian Analysis..... 171
  - 4.7 Testing Multiple Hypotheses: Variable Selection ..... 178
  - 4.8 Approaches to Variable Selection and Modeling ..... 179
    - 4.8.1 Stepwise Methods ..... 181
    - 4.8.2 All Possible Subsets..... 183
    - 4.8.3 Bayesian Model Averaging ..... 185
    - 4.8.4 Shrinkage Methods..... 185
  - 4.9 Model Building Uncertainty..... 185
  - 4.10 A Pragmatic Compromise to Variable Selection ..... 188
  - 4.11 Concluding Comments ..... 189
  - 4.12 Bibliographic Notes..... 190
  - 4.13 Exercises ..... 190

**Part II Independent Data**

- 5 Linear Models..... 195**
  - 5.1 Introduction ..... 195
  - 5.2 Motivating Example: Prostate Cancer ..... 195
  - 5.3 Model Specification..... 196
  - 5.4 A Justification for Linear Modeling..... 198
  - 5.5 Parameter Interpretation ..... 199
    - 5.5.1 Causation Versus Association ..... 199
    - 5.5.2 Multiple Parameters ..... 201
    - 5.5.3 Data Transformations ..... 205
  - 5.6 Frequentist Inference ..... 209
    - 5.6.1 Likelihood ..... 209
    - 5.6.2 Least Squares Estimation ..... 214

- 5.6.3 The Gauss–Markov Theorem ..... 215
- 5.6.4 Sandwich Estimation ..... 216
- 5.7 Bayesian Inference ..... 221
- 5.8 Analysis of Variance ..... 224
  - 5.8.1 One-Way ANOVA ..... 224
  - 5.8.2 Crossed Designs ..... 227
  - 5.8.3 Nested Designs ..... 229
  - 5.8.4 Random and Mixed Effects Models ..... 230
- 5.9 Bias-Variance Trade-Off ..... 231
- 5.10 Robustness to Assumptions ..... 236
  - 5.10.1 Distribution of Errors ..... 237
  - 5.10.2 Nonconstant Variance ..... 237
  - 5.10.3 Correlated Errors ..... 238
- 5.11 Assessment of Assumptions ..... 239
  - 5.11.1 Review of Assumptions ..... 239
  - 5.11.2 Residuals and Influence ..... 240
  - 5.11.3 Using the Residuals ..... 243
- 5.12 Example: Prostate Cancer ..... 245
- 5.13 Concluding Remarks ..... 247
- 5.14 Bibliographic Notes ..... 248
- 5.15 Exercises ..... 249
- 6 General Regression Models ..... 253**
  - 6.1 Introduction ..... 253
  - 6.2 Motivating Example: Pharmacokinetics of Theophylline ..... 254
  - 6.3 Generalized Linear Models ..... 256
  - 6.4 Parameter Interpretation ..... 259
  - 6.5 Likelihood Inference for GLMs ..... 260
    - 6.5.1 Estimation ..... 260
    - 6.5.2 Computation ..... 263
    - 6.5.3 Hypothesis Testing ..... 267
  - 6.6 Quasi-likelihood Inference for GLMs ..... 270
  - 6.7 Sandwich Estimation for GLMs ..... 272
  - 6.8 Bayesian Inference for GLMs ..... 273
    - 6.8.1 Prior Specification ..... 273
    - 6.8.2 Computation ..... 274
    - 6.8.3 Hypothesis Testing ..... 275
    - 6.8.4 Overdispersed GLMs ..... 276
  - 6.9 Assessment of Assumptions for GLMs ..... 278
  - 6.10 Nonlinear Regression Models ..... 283
  - 6.11 Identifiability ..... 284
  - 6.12 Likelihood Inference for Nonlinear Models ..... 285
    - 6.12.1 Estimation ..... 285
    - 6.12.2 Hypothesis Testing ..... 287
  - 6.13 Least Squares Inference ..... 289
  - 6.14 Sandwich Estimation for Nonlinear Models ..... 290

- 6.15 The Geometry of Least Squares ..... 291
- 6.16 Bayesian Inference for Nonlinear Models ..... 294
  - 6.16.1 Prior Specification ..... 294
  - 6.16.2 Computation ..... 294
  - 6.16.3 Hypothesis Testing ..... 295
- 6.17 Assessment of Assumptions for Nonlinear Models ..... 298
- 6.18 Concluding Remarks ..... 299
- 6.19 Bibliographic Notes ..... 299
- 6.20 Exercises ..... 300
- 7 Binary Data Models ..... 305**
  - 7.1 Introduction ..... 305
  - 7.2 Motivating Examples ..... 306
    - 7.2.1 Outcome After Head Injury ..... 306
    - 7.2.2 Aircraft Fasteners ..... 306
    - 7.2.3 Bronchopulmonary Dysplasia ..... 307
  - 7.3 The Binomial Distribution ..... 308
    - 7.3.1 Genesis ..... 308
    - 7.3.2 Rare Events ..... 309
  - 7.4 Generalized Linear Models for Binary Data ..... 310
    - 7.4.1 Formulation ..... 310
    - 7.4.2 Link Functions ..... 312
  - 7.5 Overdispersion ..... 313
  - 7.6 Logistic Regression Models ..... 316
    - 7.6.1 Parameter Interpretation ..... 316
    - 7.6.2 Likelihood Inference for Logistic Regression Models .. 318
    - 7.6.3 Quasi-likelihood Inference for Logistic Regression Models ..... 321
    - 7.6.4 Bayesian Inference for Logistic Regression Models .... 321
  - 7.7 Conditional Likelihood Inference ..... 327
  - 7.8 Assessment of Assumptions ..... 331
  - 7.9 Bias, Variance, and Collapsibility ..... 334
  - 7.10 Case-Control Studies ..... 337
    - 7.10.1 The Epidemiological Context ..... 337
    - 7.10.2 Estimation for a Case-Control Study ..... 338
    - 7.10.3 Estimation for a Matched Case-Control Study ..... 341
  - 7.11 Concluding Remarks ..... 343
  - 7.12 Bibliographic Notes ..... 344
  - 7.13 Exercises ..... 345

**Part III Dependent Data**

- 8 Linear Models ..... 353**
  - 8.1 Introduction ..... 353
  - 8.2 Motivating Example: Dental Growth Curves ..... 354

8.3	The Efficiency of Longitudinal Designs .....	356
8.4	Linear Mixed Models .....	359
	8.4.1 The General Framework .....	359
	8.4.2 Covariance Models for Clustered Data .....	360
	8.4.3 Parameter Interpretation for Linear Mixed Models.....	363
8.5	Likelihood Inference for Linear Mixed Models .....	364
	8.5.1 Inference for Fixed Effects .....	365
	8.5.2 Inference for Variance Components via Maximum Likelihood .....	367
	8.5.3 Inference for Variance Components via Restricted Maximum Likelihood .....	368
	8.5.4 Inference for Random Effects .....	376
8.6	Bayesian Inference for Linear Mixed Models .....	381
	8.6.1 A Three-Stage Hierarchical Model.....	381
	8.6.2 Hyperpriors .....	382
	8.6.3 Implementation .....	386
	8.6.4 Extensions .....	388
8.7	Generalized Estimating Equations .....	391
	8.7.1 Motivation .....	391
	8.7.2 The GEE Algorithm.....	392
	8.7.3 Estimation of Variance Parameters .....	395
8.8	Assessment of Assumptions.....	400
	8.8.1 Review of Assumptions .....	400
	8.8.2 Approaches to Assessment .....	402
8.9	Cohort and Longitudinal Effects .....	413
8.10	Concluding Remarks.....	416
8.11	Bibliographic Notes.....	416
8.12	Exercises .....	417
<b>9</b>	<b>General Regression Models.....</b>	<b>425</b>
9.1	Introduction .....	425
9.2	Motivating Examples .....	426
	9.2.1 Contraception Data.....	426
	9.2.2 Seizure Data .....	427
	9.2.3 Pharmacokinetics of Theophylline .....	428
9.3	Generalized Linear Mixed Models .....	430
9.4	Likelihood Inference for Generalized Linear Mixed Models .....	432
9.5	Conditional Likelihood Inference for Generalized Linear Mixed Models .....	437
9.6	Bayesian Inference for Generalized Linear Mixed Models .....	441
	9.6.1 Model Formulation.....	441
	9.6.2 Hyperpriors .....	441
9.7	Generalized Linear Mixed Models with Spatial Dependence ....	445
	9.7.1 A Markov Random Field Prior .....	445
	9.7.2 Hyperpriors .....	447

- 9.8 Conjugate Random Effects Models ..... 450
- 9.9 Generalized Estimating Equations for Generalized Linear Models ..... 451
- 9.10 GEE2: Connected Estimating Equations ..... 452
- 9.11 Interpretation of Marginal and Conditional Regression Coefficients ..... 455
- 9.12 Introduction to Modeling Dependent Binary Data ..... 457
- 9.13 Mixed Models for Binary Data ..... 458
  - 9.13.1 Generalized Linear Mixed Models for Binary Data ..... 458
  - 9.13.2 Likelihood Inference for the Binary Mixed Model ..... 462
  - 9.13.3 Bayesian Inference for the Binary Mixed Model ..... 462
  - 9.13.4 Conditional Likelihood Inference for Binary Mixed Models ..... 465
- 9.14 Marginal Models for Dependent Binary Data ..... 467
  - 9.14.1 Generalized Estimating Equations ..... 467
  - 9.14.2 Loglinear Models ..... 468
  - 9.14.3 Further Multivariate Binary Models ..... 471
- 9.15 Nonlinear Mixed Models ..... 475
- 9.16 Parameterization of the Nonlinear Model ..... 477
- 9.17 Likelihood Inference for the Nonlinear Mixed Model ..... 479
- 9.18 Bayesian Inference for the Nonlinear Mixed Model ..... 482
  - 9.18.1 Hyperpriors ..... 482
  - 9.18.2 Inference for Functions of Interest ..... 484
- 9.19 Generalized Estimating Equations ..... 487
- 9.20 Assessment of Assumptions for General Regression Models .... 489
- 9.21 Concluding Remarks ..... 492
- 9.22 Bibliographic Notes ..... 495
- 9.23 Exercises ..... 496

**Part IV Nonparametric Modeling**

- 10 Preliminaries for Nonparametric Regression ..... 503**
  - 10.1 Introduction ..... 503
  - 10.2 Motivating Examples ..... 504
    - 10.2.1 Light Detection and Ranging ..... 505
    - 10.2.2 Ethanol Data ..... 505
  - 10.3 The Optimal Prediction ..... 506
    - 10.3.1 Continuous Responses ..... 507
    - 10.3.2 Discrete Responses with  $K$  Categories ..... 508
    - 10.3.3 General Responses ..... 510
    - 10.3.4 In Practice ..... 511
  - 10.4 Measures of Predictive Accuracy ..... 511
    - 10.4.1 Continuous Responses ..... 512
    - 10.4.2 Discrete Responses with  $K$  Categories ..... 515
    - 10.4.3 General Responses ..... 517

10.5	A First Look at Shrinkage Methods .....	517
10.5.1	Ridge Regression .....	517
10.5.2	The Lasso .....	523
10.6	Smoothing Parameter Selection .....	526
10.6.1	Mallows $C_P$ .....	527
10.6.2	$K$ -Fold Cross-Validation .....	529
10.6.3	Generalized Cross-Validation .....	532
10.6.4	AIC for General Models .....	534
10.6.5	Cross-Validation for Generalized Linear Models .....	538
10.7	Concluding Comments .....	542
10.8	Bibliographic Notes .....	543
10.9	Exercises .....	543
<b>11</b>	<b>Spline and Kernel Methods .....</b>	<b>547</b>
11.1	Introduction .....	547
11.2	Spline Methods .....	547
11.2.1	Piecewise Polynomials and Splines .....	547
11.2.2	Natural Cubic Splines .....	552
11.2.3	Cubic Smoothing Splines .....	553
11.2.4	$B$ -Splines .....	556
11.2.5	Penalized Regression Splines .....	557
11.2.6	A Brief Spline Summary .....	560
11.2.7	Inference for Linear Smoothers .....	560
11.2.8	Linear Mixed Model Spline Representation: Likelihood Inference .....	563
11.2.9	Linear Mixed Model Spline Representation: Bayesian Inference .....	567
11.3	Kernel Methods .....	572
11.3.1	Kernels .....	574
11.3.2	Kernel Density Estimation .....	575
11.3.3	The Nadaraya–Watson Kernel Estimator .....	578
11.3.4	Local Polynomial Regression .....	580
11.4	Variance Estimation .....	584
11.5	Spline and Kernel Methods for Generalized Linear Models .....	587
11.5.1	Generalized Linear Models with Penalized Regression Splines .....	587
11.5.2	A Generalized Linear Mixed Model Spline Representation .....	591
11.5.3	Generalized Linear Models with Local Polynomials .....	592
11.6	Concluding Comments .....	593
11.7	Bibliographic Notes .....	593
11.8	Exercises .....	594

**12 Nonparametric Regression with Multiple Predictors** ..... 597

    12.1 Introduction ..... 597

    12.2 Generalized Additive Models ..... 598

        12.2.1 Model Formulation ..... 598

        12.2.2 Computation via Backfitting ..... 599

    12.3 Spline Methods with Multiple Predictors ..... 601

        12.3.1 Natural Thin Plate Splines ..... 602

        12.3.2 Thin Plate Regression Splines ..... 603

        12.3.3 Tensor Product Splines ..... 604

    12.4 Kernel Methods with Multiple Predictors ..... 607

    12.5 Smoothing Parameter Estimation ..... 608

        12.5.1 Conventional Approaches ..... 608

        12.5.2 Mixed Model Formulation ..... 608

    12.6 Varying-Coefficient Models ..... 610

    12.7 Regression Trees ..... 614

        12.7.1 Hierarchical Partitioning ..... 614

        12.7.2 Multiple Adaptive Regression Splines ..... 622

    12.8 Classification ..... 624

        12.8.1 Logistic Models with  $K$  Classes ..... 625

        12.8.2 Linear and Quadratic Discriminant Analysis ..... 626

        12.8.3 Kernel Density Estimation and Classification ..... 630

        12.8.4 Classification Trees ..... 634

        12.8.5 Bagging ..... 636

        12.8.6 Random Forests ..... 639

    12.9 Concluding Comments ..... 643

    12.10 Bibliographic Notes ..... 644

    12.11 Exercises ..... 644

**Part V Appendices**

**A Differentiation of Matrix Expressions** ..... 649

**B Matrix Results** ..... 653

**C Some Linear Algebra** ..... 655

**D Probability Distributions and Generating Functions** ..... 657

**E Functions of Normal Random Variables** ..... 667

**F Some Results from Classical Statistics** ..... 669

**G Basic Large Sample Theory** ..... 673

**References** ..... 675

**Index** ..... 689

