Vector Generalized Linear And Additive Models With An Implementation In R Springer Series In Statistics

This easy-to-follow applied book on semiparametric regression methods using R is intended to close the gap between the available methodology and its use in practice. Semiparametric regression has a large literature but much of it is geared towards data analysts who have advanced knowledge of statistical methods. While R now has a great deal of semiparametric regression functionality, many of these developments have not trickled down to rank-and-file statistical analysts. The authors assemble a broad range of semiparametric regression R analyses and put them in a form that is useful for applied researchers. There are chapters devoted to penalized spines, generalized additive models, grouped data, bivariate extensions of penalized spines, and spatial semi-parametric regression models. Where feasible, the R code is provided in the text, however the book is also accompanied by an external website complete with datasets and R code. Because of its flexibility, semiparametric regression has proven to be of great value with many applications in fields as diverse as astronomy, biology, medicine, economics, and finance. This book is intended for applied statistical analysts who have some familiarity with R.

Flexible Bayesian Regression Modeling is a step-by-step guide to the Bayesian revolution in regression modeling, for use in advanced econometric and statistical analysis where datasets are characterized by complexity, multiplicity, and large sample sizes, necessitating the need for considerable flexibility in modeling techniques. It reviews three forms of flexibility: methods which provide flexibility in their error distribution; methods which model non-central parts of the distribution (such as quantile regression); and finally models that allow the mean function to be flexible (such as spline models). Each chapter discusses the key aspects of fitting a regression model. R programs accompany the methods. This book is particularly relevant to non-specialist practitioners with intermediate mathematical training seeking to apply Bayesian approaches in economics, biology, finance, engineering and medicine. Introduces powerful new nonparametric Bayesian regression techniques to classically trained practitioners Focuses on approaches offering both superior power and methodological flexibility Supplemented with instructive and relevant R programs within the text Covers linear regression, nonlinear regression and quantile regression techniques Provides diverse disciplinary case studies for correlation and optimization problems drawn from Bayesian analysis 'in the wild'

The essential introduction to the theory and application of linear models—now in a valuable new edition Since most advanced statistical tools are generalizations of the linear model, it is neces-sary to first master the linear model in order

to move forward to more advanced concepts. The linear model remains the main tool of the applied statistician and is central to the training of any statistician regardless of whether the focus is applied or theoretical. This completely revised and updated new edition successfully develops the basic theory of linear models for regression, analysis of variance, analysis of covariance, and linear mixed models. Recent advances in the methodology related to linear mixed models, generalized linear models, and the Bayesian linear model are also addressed. Linear Models in Statistics, Second Edition includes full coverage of advanced topics, such as mixed and generalized linear models, Bayesian linear models, two-way models with empty cells, geometry of least squares, vector-matrix calculus, simultaneous inference, and logistic and nonlinear regression. Algebraic, geometrical, frequentist, and Bayesian approaches to both the inference of linear models and the analysis of variance are also illustrated. Through the expansion of relevant material and the inclusion of the latest technological developments in the field, this book provides readers with the theoretical foundation to correctly interpret computer software output as well as effectively use, customize, and understand linear models. This modern Second Edition features: New chapters on Bayesian linear models as well as random and mixed linear models Expanded discussion of two-way models with empty cells Additional sections on the geometry of least squares Updated coverage of simultaneous inference The book is complemented with easy-to-read proofs, real data sets, and an extensive bibliography. A thorough review of the requisite matrix algebra has been addedfor transitional purposes, and numerous theoretical and applied problems have been incorporated with selected answers provided at the end of the book. A related Web site includes additional data sets and SAS® code for all numerical examples. Linear Model in Statistics, Second Edition is a must-have book for courses in statistics, biostatistics, and mathematics at the upper-undergraduate and graduate levels. It is also an invaluable reference for researchers who need to gain a better understanding of regression and analysis of variance.

The high-level language of R is recognized as one of the most powerful and flexible statistical software environments, and is rapidly becoming the standard setting for quantitative analysis, statistics and graphics. R provides free access to unrivalled coverage and cutting-edge applications, enabling the user to apply numerous statistical methods ranging from simple regression to time series or multivariate analysis. Building on the success of the author's bestselling Statistics: An Introduction using R, The R Book is packed with worked examples, providing an all inclusive guide to R, ideal for novice and more accomplished users alike. The book assumes no background in statistics or computing and introduces the advantages of the R environment, detailing its applications in a wide range of disciplines. Provides the first comprehensive reference manual for the R language, including practical guidance and full coverage of the graphics facilities. Introduces all the statistical models covered by R, beginning with simple classical tests such as chi-square and t-

test. Proceeds to examine more advance methods, from regression and analysis of variance, through to generalized linear models, generalized mixed models, time series, spatial statistics, multivariate statistics and much more. The R Book is aimed at undergraduates, postgraduates and professionals in science, engineering and medicine. It is also ideal for students and professionals in statistics, economics, geography and the social sciences.

This book was written to rigorously illustrate the practical application of the projective approach to linear models. To some, this may seem contradictory. I contend that it is possible to be both rigorous and illustrative and that it is possible to use the projective approach in practical applications. Therefore, unlike many other books on linear models, the use of projections and sub spaces does not stop after the general theory. They are used wherever I could figure out how to do it. Solving normal equations and using calculus (outside of maximum likelihood theory) are anathema to me. This is because I do not believe that they contribute to the understanding of linear models. I have similar feelings about the use of side conditions. Such topics are mentioned when appropriate and thenceforward avoided like the plague. On the other side of the coin, I just as strenuously reject teaching linear models with a coordinate free approach. Although Joe Eaton assures me that the issues in complicated problems frequently become clearer when considered free of coordinate systems, my experience is that too many people never make the jump from coordinate free theory back to practical applications. I think that coordinate free theory is better tackled after mastering linear models from some other approach. In particular, I think it would be very easy to pick up the coordinate free approach after learning the material in this book. See Eaton (1983) for an excellent exposition of the coordinate free approach.

A groundbreaking introduction to vectors, matrices, and least squares for engineering applications, offering a wealth of practical examples.

This book is the collection of the Abstract / Short Papers submitted by the authors of the International Conference of The CLAssification and Data Analysis Group (CLADAG) of the Italian Statistical Society (SIS), held in Milan (Italy) on September 13-15, 2017.

This is the only book actuaries need to understand generalized linear models (GLMs) for insurance applications. GLMs are used in the insurance industry to support critical decisions. Until now, no text has introduced GLMs in this context or addressed the problems specific to insurance data. Using insurance data sets, this practical, rigorous book treats GLMs, covers all standard exponential family distributions, extends the methodology to correlated data structures, and discusses recent developments which go beyond the GLM. The issues in the book are specific to insurance data, such as model selection in the presence of large data sets and the handling of varying exposure times. Exercises and data-based practicals help readers to consolidate their skills, with solutions and data sets given on the companion website. Although

the book is package-independent, SAS code and output examples feature in an appendix and on the website. In addition, R code and output for all the examples are provided on the website.

This book offers a detailed application guide to XploRe - an interactive statistical computing environment. As a guide it contains case studies of real data analysis situations. It helps the beginner in statistical data analysis to learn how XploRe works in real life applications. Many examples from practice are discussed and analysed in full length. Great emphasis is put on a graphic based understanding of the data interrelations. The case studies include: Survival modelling with Cox's proportional hazard regression, Vitamin C data analysis with Quantile Regression, and many others.

This COMPSTAT 2002 book contains the Keynote, Invited, and Full Contributed papers presented in Berlin, August 2002. A companion volume including Short Communications and Posters is published on CD. The COMPSTAT 2002 is the 15th conference in a serie of biannual conferences with the objective to present the latest developments in Computational Statistics and is taking place from August 24th to August 28th, 2002. Previous COMPSTATs were in Vienna (1974), Berlin (1976), Leiden (1978), Edinburgh (1980), Toulouse (1982), Pra~ue (1984), Rome (1986), Copenhagen (1988), Dubrovnik (1990), Neuchatel (1992), Vienna (1994), Barcelona (1996), Bris tol (1998) and Utrecht (2000). COMPSTAT 2002 is organised by CASE, Center of Applied Statistics and Eco nomics at Humboldt-Universitat zu Berlin in cooperation with F'reie Universitat Berlin and University of Potsdam. The topics of COMPSTAT include methodological applications, innovative soft ware and mathematical developments, especially in the following fields: statistical risk management, multivariate and robust analysis, Markov Chain Monte Carlo Methods, statistics of E-commerce, new strategies in teaching (Multimedia, In ternet), computerbased sampling/questionnaires, analysis of large databases (with emphasis on computing in memory), graphical tools for data analysis, classification and clustering, new statistical software and historical development of software.

Vector generalized additive models (VGAMs) are an extension of the class of generalized additive models (GAMs) to include multivariate regression models in a very natural way by using vector smoothing. The current VGAM class is very large and includes many statistical distributions and models, for example, univariate and multivariate distributions, categorical data analysis, quantile and expectile regression, time series, survival analysis, extreme value analysis, and nonlinear least-squares models. Parameter estimation is performed by a combination of IRLS and modified vector backfitting using vector splines. A major issue, however, is that it is not easy to efficiently integrate smoothness estimation methods with the backfitting approach. The aim of this research study is to introduce a new efficient method based on penalized regression splines for estimating parameter coefficients to the VGAM class, and to integrate automatic numerical procedures to determine the shape of non-linear terms from the data into the VGAM framework. To achieve these, we develop VGAMs based on penalized regression splines using P-spline smoothers, which we term 'P-spline VGAMs'. P-spline VGAMs are represented in this thesis as penalized vector generalized linear models (VGLMs), where each smooth component of a P-spline VGAM is represented using penalized B-splines or P-spline smoothers and has an associated discrete penalty measuring its wiggliness controlled by the smoothing parameter. P-spline VGAMs can be then fitted by the usual iteratively reweighted least squares (IRLS) scheme for VGLMs, except that a penalized least squares problem, in which the set of smoothing parameters must be estimated alongside the other model parameters, is solved at each iterate. The smoothing parameters are estimated by minimizing the approximate unbiased risk estimator (UBRE) using the computational procedure for the automatic and stable multiple smoothing parameter selection based on the pivoted QR decomposition and singular value decomposition.

Importantly, the new fitting procedure is developed for the full range of VGAM models involving infrastructure such as constraints on model terms. This research study describes the theoretical and practical aspects of the proposed method (Pspline vector generalized additive models). The methods have been implemented as R functions and the practical performance of the proposed method is investigated and compared to the existing approaches (VGAMs based on the classical backfitting) via simulation. As an illustration of the developments, the proposed method is applied to data from a cross-sectional workforce study combined with a health survey from New Zealand during the 1990s, and data from a survey study of the pregnancy and birth process during 1990 2004; using several statistical models, which include the multinomial logit, proportional and non-proportional odds models, bivariate logistic model, and the LMS method for quantile regression.

A valuable overview of the most important ideas and results in statistical modeling Written by a highly-experienced author, Foundations of Linear and Generalized Linear Models is a clear and comprehensive guide to the key concepts and results of linearstatistical models. The book presents a broad, in-depth overview of the most commonly usedstatistical models by discussing the theory underlying the models, R software applications,and examples with crafted models to elucidate key ideas and promote practical modelbuilding. The book begins by illustrating the fundamentals of linear models, such as how the model-fitting projects the data onto a model vector subspace and how orthogonal decompositions of the data yield information about the effects of explanatory variables. Subsequently, the book covers the most popular generalized linear models, which include binomial and multinomial logistic regression for categorical data, and Poisson and negative binomial loglinear models for count data. Focusing on the theoretical underpinnings of these models, Foundations ofLinear and Generalized Linear Models also features: An introduction to quasi-likelihood methods that require weaker distributional assumptions, such as generalized estimating equation methods An overview of linear mixed models and generalized linear mixed models with random effects for clustered correlated data, Bayesian modeling, and extensions to handle problematic cases such as high dimensional problems Numerous examples that use R software for all text data analyses More than 400 exercises for readers to practice and extend the theory, methods, and data analysis A supplementary website with datasets for the examples and exercises An invaluable textbook for upper-undergraduate and graduate-level students in statistics and biostatistics courses, Foundations of Linear and Generalized Linear Models is also an excellent reference for practicing statisticians and biostatisticians, as well as anyone who is interested in learning about the most important statistical models for analyzing data.

State space models have gained tremendous popularity in recent years in as disparate fields as engineering, economics, genetics and ecology. After a detailed introduction to general state space models, this book focuses on dynamic linear models, emphasizing their Bayesian analysis. Whenever possible it is shown how to compute estimates and forecasts in closed form; for more complex models, simulation techniques are used. A final chapter covers modern sequential Monte Carlo algorithms. The book illustrates all the fundamental steps needed to use dynamic linear models in practice, using R. Many detailed examples based on real data sets are provided to show how to set up a specific model, estimate its parameters, and use it for forecasting. All the code used in the book is available online. No prior knowledge of Bayesian statistics or time series analysis is required, although familiarity with basic statistics and R is assumed.

Although standard mixed effects models are useful in a range of studies, other approaches must often be used in correlation with them when studying complex or incomplete data. Mixed Effects Models for Complex Data discusses commonly used mixed effects models and presents appropriate approaches to address dropouts, missing data, measurement errors, censoring, and outliers. For each class of mixed effects model, the author reviews the corresponding class of regression model for cross-sectional data. An overview of general models and methods,

along with motivating examples After presenting real data examples and outlining general approaches to the analysis of longitudinal/clustered data and incomplete data, the book introduces linear mixed effects (LME) models, generalized linear mixed models (GLMMs), nonlinear mixed effects (NLME) models, and semiparametric and nonparametric mixed effects models. It also includes general approaches for the analysis of complex data with missing values, measurement errors, censoring, and outliers. Self-contained coverage of specific topics Subsequent chapters delve more deeply into missing data problems, covariate measurement errors, and censored responses in mixed effects models. Focusing on incomplete data, the book also covers survival and frailty models, joint models of survival and longitudinal data, robust methods for mixed effects models, marginal generalized estimating equation (GEE) models for longitudinal or clustered data, and Bayesian methods for mixed effects models. Background material In the appendix, the author provides background information, such as likelihood theory, the Gibbs sampler, rejection and importance sampling methods, numerical integration methods, optimization methods, bootstrap, and matrix algebra. Failure to properly address missing data, measurement errors, and other issues in statistical analyses can lead to severely biased or misleading results. This book explores the biases that arise when naïve methods are used and shows which approaches should be used to achieve accurate results in longitudinal data analysis.

Discover New Methods for Dealing with High-Dimensional Data A sparse statistical model has only a small number of nonzero parameters or weights; therefore, it is much easier to estimate and interpret than a dense model. Statistical Learning with Sparsity: The Lasso and Generalizations presents methods that exploit sparsity to help recover the underlying signal in a set of data. Top experts in this rapidly evolving field, the authors describe the lasso for linear regression and a simple coordinate descent algorithm for its computation. They discuss the application of l1 penalties to generalized linear models and support vector machines, cover generalized penalties such as the elastic net and group lasso, and review numerical methods for optimization. They also present statistical inference methods for fitted (lasso) models, including the bootstrap, Bayesian methods, and recently developed approaches. In addition, the book examines matrix decomposition, sparse multivariate analysis, graphical models, and compressed sensing. It concludes with a survey of theoretical results for the lasso. In this age of big data, the number of features measured on a person or object can be large and might be larger than the number of observations. This book shows how the sparsity assumption allows us to tackle these problems and extract useful and reproducible patterns from big datasets. Data analysts, computer scientists, and theorists will appreciate this thorough and up-to-date treatment of sparse statistical modeling.

Now in widespread use, generalized additive models (GAMs) have evolved into a standard statistical methodology of considerable flexibility. While Hastie and Tibshirani's outstanding 1990 research monograph on GAMs is largely responsible for this, there has been a long-standing need for an accessible introductory treatment of the subject that also emphasizes recent penalized regression spline approaches to GAMs and the mixed model extensions of these models. Generalized Additive Models: An Introduction with R imparts a thorough understanding of the theory and practical applications of GAMs and related advanced models, enabling informed use of these very flexible tools. The author bases his approach on a framework of penalized regression splines, and builds a well-grounded foundation through motivating chapters on linear and generalized linear models. While firmly focused on the practical aspects of GAMs, discussions include fairly full explanations of the theory underlying the methods. Use of the freely available R software helps explain the theory and illustrates the practicalities of linear, generalized linear, and generalized additive models, as well as their mixed effect extensions. The treatment is rich with practical examples, and it includes an entire chapter on the analysis of real data sets using R and the author's add-on package mgcv. Each chapter includes

exercises, for which complete solutions are provided in an appendix. Concise, comprehensive, and essentially self-contained, Generalized Additive Models: An Introduction with R prepares readers with the practical skills and the theoretical background needed to use and understand GAMs and to move on to other GAM-related methods and models, such as SS-ANOVA, P-splines, backfitting and Bayesian approaches to smoothing and additive modelling.

This book is about learning from data using the Generalized Additive Models for Location, Scale and Shape (GAMLSS). GAMLSS extends the Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs) to accommodate large complex datasets, which are increasingly prevalent. In particular, the GAMLSS statistical framework enables flexible regression and smoothing models to be fitted to the data. The GAMLSS model assumes that the response variable has any parametric (continuous, discrete or mixed) distribution which might be heavy- or light-tailed, and positively or negatively skewed. In addition, all the parameters of the distribution (location, scale, shape) can be modelled as linear or smooth functions of explanatory variables. Key Features: Provides a broad overview of flexible regression and smoothing techniques to learn from data whilst also focusing on the practical application of methodology using GAMLSS software in R. Includes a comprehensive collection of real data examples, which reflect the range of problems addressed by GAMLSS models and provide a practical illustration of the process of using flexible GAMLSS models for statistical learning. R code integrated into the text for ease of understanding and replication. Supplemented by a website with code, data and extra materials. This book aims to help readers understand how to learn from data encountered in many fields. It will be useful for practitioners and researchers who wish to understand and use the GAMLSS models to learn from data and also for students who wish to learn GAMLSS through practical examples.

An Introduction to Statistical Learning provides an accessible overview of the field of statistical learning, an essential toolset for making sense of the vast and complex data sets that have emerged in fields ranging from biology to finance to marketing to astrophysics in the past twenty years. This book presents some of the most important modeling and prediction techniques, along with relevant applications. Topics include linear regression, classification, resampling methods, shrinkage approaches, tree-based methods, support vector machines, clustering, and more. Color graphics and real-world examples are used to illustrate the methods presented. Since the goal of this textbook is to facilitate the use of these statistical learning techniques by practitioners in science, industry, and other fields, each chapter contains a tutorial on implementing the analyses and methods presented in R, an extremely popular open source statistical software platform. Two of the authors co-wrote The Elements of Statistical Learning (Hastie, Tibshirani and Friedman, 2nd edition 2009), a popular reference book for statistics and machine learning researchers. An Introduction to Statistical Learning covers many of the same topics, but at a level accessible to a much broader audience. This book is targeted at statisticians and non-statisticians alike who wish to use cutting-edge statistical learning techniques to analyze their data. The text assumes only a previous course in linear regression and no knowledge of matrix algebra.

This book presents a greatly enlarged statistical framework compared to generalized linear models (GLMs) with which to approach regression modelling. Comprising of about half-a-dozen major classes of statistical models, and fortified with necessary

infrastructure to make the models more fully operable, the framework allows analyses based on many semi-traditional applied statistics models to be performed as a coherent whole. Since their advent in 1972, GLMs have unified important distributions under a single umbrella with enormous implications. However, GLMs are not flexible enough to cope with the demands of practical data analysis. And data-driven GLMs, in the form of generalized additive models (GAMs), are also largely confined to the exponential family. The methodology here and accompanying software (the extensive VGAM R package) are directed at these limitations and are described comprehensively for the first time in one volume. This book treats distributions and classical models as generalized regression models, and the result is a much broader application base for GLMs and GAMs. The book can be used in senior undergraduate or first-year postgraduate courses on GLMs or categorical data analysis and as a methodology resource for VGAM users. In the second part of the book, the R package VGAM allows readers to grasp immediately applications of the methodology. R code is integrated in the text, and datasets are used throughout. Potential applications include ecology, finance, biostatistics, and social sciences. The methodological contribution of this book stands alone and does not require use of the VGAM package.

This textbook presents an introduction to multiple linear regression, providing real-world data sets and practice problems. A practical working knowledge of applied statistical practice is developed through the use of these data sets and numerous case studies. The authors include a set of practice problems both at the end of each chapter and at the end of the book. Each example in the text is cross-referenced with the relevant data set, so that readers can load the data and follow the analysis in their own R sessions. The balance between theory and practice is evident in the list of problems, which vary in difficulty and purpose. This book is designed with teaching and learning in mind, featuring chapter introductions and summaries, exercises, short answers, and simple, clear examples. Focusing on the connections between generalized linear models (GLMs) and linear regression, the book also references advanced topics and tools that have not typically been included in introductions to GLMs to date, such as Tweedie family distributions with power variance functions, saddlepoint approximations, likelihood score tests, modified profile likelihood, and randomized quantile residuals. In addition, the authors introduce the new R code package, GLMsData, created specifically for this book. Generalized Linear Models with Examples in R balances theory with practice, making it ideal for both introductory and graduate-level students who have a basic knowledge of matrix algebra, calculus, and statistics.

This book describes how generalised linear modelling procedures can be used in many different fields, without becoming entangled in problems of statistical inference. The author shows the unity of many of the commonly used models and provides readers with a taste of many different areas, such as survival models, time series, and spatial analysis, and of their unity. As such, this book will appeal to applied statisticians and to scientists having a basic grounding in modern statistics. With many exercises at the end of each chapter, it will equally constitute an excellent text for teaching applied statistics students and non- statistics majors. The reader is assumed to have knowledge of basic statistical principles, whether from a Bayesian, frequentist, or direct likelihood point of view, being familiar at least with the analysis of the simpler normal linear models, regression and ANOVA.

The first edition of this book has established itself as one of the leading references on generalized additive models (GAMs), and the only book on the topic to be introductory in nature with a wealth of practical examples and software implementation. It is self-contained, providing the necessary background in linear models, linear mixed models, and generalized linear models (GLMs), before presenting a balanced treatment of the theory and applications of GAMs and related models. The author bases his approach on a framework of penalized regression splines, and while firmly focused on the practical aspects of GAMs, discussions include fairly full explanations of the theory underlying the methods. Use of R software helps explain the theory and illustrates the practical application of the methodology. Each chapter contains an extensive set of exercises, with solutions in an appendix or in the book's R data package gamair, to enable use as a course text or for self-study. Simon N. Wood is a professor of Statistical Science at the University of Bristol, UK, and author of the R package mgcv. Combining a modern, data-analytic perspective with a focus on applications in the social sciences, the Third Edition of Applied Regression Analysis and Generalized Linear Models provides in-depth coverage of regression analysis, generalized linear models, and closely related methods, such as bootstrapping and missing data. Updated throughout, this Third Edition includes new chapters on mixed-effects models for hierarchical and longitudinal data. Although the text is largely accessible to readers with a modest background in statistics and mathematics, author John Fox also presents more advanced material in optional sections and chapters throughout the book. Available with Perusall—an eBook that makes it easier to prepare for class Perusall is an award-winning eBook platform featuring social annotation tools that allow students and instructors to collaboratively mark up and discuss their SAGE textbook. Backed by research and supported by technological innovations developed at Harvard University, this process of learning through collaborative annotation keeps your students engaged and makes teaching easier and more effective. Learn more. Statistical Regression and Classification: From Linear Models to Machine Learning takes an innovative look at the traditional statistical regression course, presenting a contemporary treatment in line with today's applications and users. The text takes a modern look at regression: * A thorough treatment of classical linear and generalized linear models, supplemented with introductory material on machine learning methods. * Since classification is the focus of many contemporary applications, the book covers this topic in detail, especially the multiclass case. * In view of the voluminous nature of many modern datasets, there is a chapter on Big Data. * Has special Mathematical and Computational Complements sections at ends of chapters, and exercises are partitioned into Data, Math and Complements problems. * Instructors can tailor coverage for specific audiences such as majors in Statistics, Computer Science, or Economics. * More than 75 examples using real data. The book treats classical regression methods in an innovative, contemporary manner. Though some statistical learning methods are introduced, the primary methodology used is linear and generalized linear parametric models, covering both the Description and Prediction goals of regression methods. The author is just as interested in Description applications of regression, such as measuring the gender wage gap in Silicon Valley, as in forecasting tomorrow's demand for bike rentals. An entire chapter is devoted to measuring such effects,

including discussion of Simpson's Paradox, multiple inference, and causation issues. Similarly, there is an entire chapter of parametric model fit, making use of both residual analysis and assessment via nonparametric analysis. Norman Matloff is a professor of computer science at the University of California, Davis, and was a founder of the Statistics Department at that institution. His current research focus is on recommender systems, and applications of regression methods to small area estimation and bias reduction in observational studies. He is on the editorial boards of the Journal of Statistical Computation and the R Journal. An award-winning teacher, he is the author of The Art of R Programming and Parallel Computation in Data Science: With Examples in R, C++ and CUDA.

This second edition of Hilbe's Negative Binomial Regression is a substantial enhancement to the popular first edition. The only text devoted entirely to the negative binomial model and its many variations, nearly every model discussed in the literature is addressed. The theoretical and distributional background of each model is discussed, together with examples of their construction, application, interpretation and evaluation. Complete Stata and R codes are provided throughout the text, with additional code (plus SAS), derivations and data provided on the book's website. Written for the practising researcher, the text begins with an examination of risk and rate ratios, and of the estimating algorithms used to model count data. The book then gives an in-depth analysis of Poisson regression and an evaluation of the meaning and nature of overdispersion, followed by a comprehensive analysis of the negative binomial distribution and of its parameterizations into various models for evaluating count data.

Elections are random events. From individuals deciding whether to vote, to people deciding for whom to vote, to election authorities deciding what to count, the outcomes of competitive democratic elections are rarely known until election day...or beyond. Understanding Elections through Statistics: Polling, Prediction, and Testing explores this random phenomenon from two points of view: predicting the election outcome using opinion polls and testing the election outcome using government-reported data. Written for those with only a brief introduction to statistics, this book takes you on a statistical journey from how polls are taken to how they can—and should—be used to estimate current popular opinion. Once an understanding of the election process is built, we turn toward testing elections for evidence of unfairness. While holding elections has become the de facto proof of government legitimacy, those electoral processes may hide a dirty little secret of the government illicitly ensuring a favorable election outcome. This book includes these features designed to make your statistical journey more enjoyable: Vignettes of elections, including maps, to provide concrete bases for the material In-chapter cues to help one avoid the heavy math—or to focus on it End-of-chapter problems designed to review and extend that which was covered in the chapter Many opportunities to turn the power of the R statistical environment to the enclosed election data files, as well as to those you find interesting From these features, it is clear the audience for this book is quite diverse. This text provides mathematics for those interested in mathematics, but also offers detours for those who just want a good read and a deeper understanding of elections. Author Ole J. Forsberg holds PhDs in both political science and statistics. He currently teaches mathematics and statistics in the Department of Mathematics at Knox College in Galesburg, IL.

Extreme Value Modeling and Risk Analysis: Methods and Applications presents a broad overview of statistical modeling of extreme events along with the most recent methodologies and various applications. The book brings together background material and advanced topics, eliminating the need to sort through the massive amount of literature on the subject. After reviewing univariate extreme value analysis and multivariate extremes, the book explains univariate extreme value mixture modeling, threshold selection in extreme value analysis, and threshold modeling of non-stationary extremes. It presents new results for block-maxima of vine copulas, develops time series of extremes with applications from climatology, describes max-autoregressive and moving maxima models for extremes, and discusses spatial extremes and max-stable processes. The book then covers simulation and conditional simulation of max-stable processes; inference methodologies, such as composite likelihood, Bayesian inference, and approximate Bayesian computation; and inferences about extreme quantiles and extreme dependence. It also explores novel applications of extreme value modeling, including financial investments, insurance and financial risk management, weather and climate disasters, clinical trials, and sports statistics. Risk analyses related to extreme events require the combined expertise of statisticians and domain experts in climatology, hydrology, finance, insurance, sports, and other fields. This book connects statistical/mathematical research with critical decision and risk assessment/management applications to stimulate more collaboration between these statisticians and specialists.

Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives presents regression models for ordinal outcomes, which are variables that have ordered categories but unknown spacing between the categories. The book provides comprehensive coverage of the three major classes of ordered regression models (cumulative, stage, and adjacent) as well as variations based on the application of the parallel regression assumption. The authors first introduce the three "parallel" ordered regression models before covering unconstrained partial, constrained partial, and nonparallel models. They then review existing tests for the parallel regression assumption, propose new variations of several tests, and discuss important practical concerns related to tests of the parallel regression assumption. The book also describes extensions of ordered regression models, including heterogeneous choice models, multilevel ordered models, and the Bayesian approach to ordered regression models. Some chapters include brief examples using Stata and R. This book offers a conceptual framework for understanding ordered regression models based on the probability of interest and the application of the parallel regression assumption. It demonstrates the usefulness of numerous modeling alternatives, showing you how to select the most appropriate model given the type of ordinal outcome and restrictiveness of the parallel assumption for each variable. Web Resource More detailed examples are available on a supplementary website. The site also contains JAGS, R, and Stata codes to estimate the models along with syntax to reproduce the results.

Data mining is the process of automatically searching large volumes of data for models and patterns using computational techniques from statistics, machine learning and information theory; it is the ideal tool for such an extraction of knowledge. Data mining is usually associated with a business or an organization's need to identify trends and profiles, allowing, for example, retailers to discover patterns on which to base marketing objectives. This book looks at both classical and recent techniques of

data mining, such as clustering, discriminant analysis, logistic regression, generalized linear models, regularized regression, PLS regression, decision trees, neural networks, support vector machines, Vapnik theory, naive Bayesian classifier, ensemble learning and detection of association rules. They are discussed along with illustrative examples throughout the book to explain the theory of these methods, as well as their strengths and limitations. Key Features: Presents a comprehensive introduction to all techniques used in data mining and statistical learning, from classical to latest techniques. Starts from basic principles up to advanced concepts. Includes many step-by-step examples with the main software (R, SAS, IBM SPSS) as well as a thorough discussion and comparison of those software. Gives practical tips for data mining implementation to solve real world problems. Looks at a range of tools and applications, such as association rules, web mining and text mining, with a special focus on credit scoring. Supported by an accompanying website hosting datasets and user analysis. Statisticians and business intelligence analysts, students as well as computer science, biology, marketing and financial risk professionals in both commercial and government organizations across all business and industry sectors will benefit from this book.
Vector Generalized Linear and Additive ModelsWith an Implementation in RSpringer
This book describes an array of power tools for data analysis that are based on nonparametric regression and smoothing techniques. These methods relax the linear assumption of many standard models and allow analysts to uncover structure in the data that might otherwise have been missed. While McCullagh and Nelder's Generalized Linear Models shows how to extend the usual linear methodology to cover analysis of a range of data types, Generalized Additive Models enhances this methodology even further by incorporating the flexibility of nonparametric regression. Clear prose, exercises in each chapter, and case studies enhance this popular text.
This is a book about statistical distributions, their properties, and their application to modelling the dependence of the location, scale, and shape of the distribution of a response variable on explanatory variables. It will be especially useful to applied statisticians and data scientists in a wide range of application areas, and also to those interested in the theoretical properties of distributions. This book follows the earlier book 'Flexible Regression and Smoothing: Using GAMLSS in R', [Stasinopoulos et al., 2017], which focused on the GAMLSS model and software. GAMLSS (the Generalized Additive Model for Location, Scale, and Shape, [Rigby and Stasinopoulos, 2005]), is a regression framework in which the response variable can have any parametric distribution and all the distribution parameters can be modelled as linear or smooth functions of explanatory variables. The current book focuses on distributions and their application. Key features: Describes over 100 distributions, (implemented in the GAMLSS packages in R), including continuous, discrete and mixed distributions. Comprehensive summary tables of the properties of the distributions. Discusses properties of distributions, including skewness, kurtosis, robustness and an important classification of tail heaviness. Includes mixed distributions which are continuous distributions with additional specific values with point probabilities. Includes many real data examples, with R code integrated in the text for ease of understanding and replication. Supplemented by the gamlss website. This book will be useful for applied statisticians and data scientists in selecting a distribution for a univariate

response variable and modelling its dependence on explanatory variables, and to those interested in the properties of distributions. Deftly balancing theory and application, this book stands out in its coverage of the derivation of the GLM families and their foremost links. This edition has new sections on discrete response models, including zero-truncated, zero-inflated, censored, and hurdle count models, as well as heterogeneous negative binomial, and more.

An Applied Treatment of Modern Graphical Methods for Analyzing Categorical DataDiscrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data presents an applied treatment of modern methods for the analysis of categorical data, both discrete response data and frequency data. It explains how to use graphical meth

This book contains the keynote, invited and full contributed papers presented at COMPSTAT 2000, held in Utrecht. The papers range over all aspects of the link between statistical theory and applied statistics, with special attention for developments in the area of official statistics. The papers have been thoroughly refereed.

This book brings the power of multivariate statistics to graduate-level practitioners, making these analytical methods accessible without lengthy mathematical derivations. Using the open source, shareware program R, Professor Zelterman demonstrates the process and outcomes for a wide array of multivariate statistical applications. Chapters cover graphical displays, linear algebra, univariate, bivariate and multivariate normal distributions, factor methods, linear regression, discrimination and classification, clustering, time series models, and additional methods. Zelterman uses practical examples from diverse disciplines to welcome readers from a variety of academic specialties. Those with backgrounds in statistics will learn new methods while they review more familiar topics. Chapters include exercises, real data sets, and R implementations. The data are interesting, real-world topics, particularly from health and biology-related contexts. As an example of the approach, the text examines a sample from the Behavior Risk Factor Surveillance System, discussing both the shortcomings of the data as well as useful analyses. The text avoids theoretical derivations beyond those needed to fully appreciate the methods. Prior experience with R is not necessary.

Copyright: f724f38cc499548eff9532e3fd6df1cf