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

This textbook presents an introduction to generalized linear models, complete with real-world data sets and practice problems, making it applicable for both beginning and advanced students of applied statistics. Generalized linear models (GLMs) are powerful tools in applied statistics that extend the ideas of multiple linear regression and analysis of variance to include response variables that are not normally distributed. As such, GLMs can model a wide variety of data types including counts, proportions, and binary outcomes or positive quantities. The book is designed with the student in mind, making it suitable for self-study or a structured course. Beginning with an introduction to linear regression, the book also devotes time to advanced topics not typically included in introductory textbooks. It features chapter introductions and summaries, clear examples, and many practice problems, all carefully designed to balance theory and practice. The text also provides a working knowledge of applied statistical practice through the extensive use of R, which is integrated into the text. Other features include: • Advanced topics such as power variance functions, saddlepoint approximations, likelihood score tests, modified profile likelihood, small-dispersion asymptotics, and randomized quantile residuals • Nearly 100 data sets in the companion R package GLMsData • Examples that are cross-referenced to the companion data set, allowing readers to load the data and follow the analysis in their own R session

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 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.

Applied Linear Statistical Models 5e is the long established leading authoritative text and reference on statistical modeling. For students in most any discipline where statistical analysis or interpretation is used, ALSM serves as the standard work. The text includes brief introductory and review material, and then proceeds through regression and modeling for the first half, and through ANOVA and Experimental Design in the second half. All topics are presented in a precise and clear style supported with solved examples, numbered formulae, graphic illustrations, and "Notes" to provide depth and statistical accuracy and precision. Applications used within the text and the hallmark problems, exercises, and projects are drawn from virtually all disciplines and fields providing motivation for students in virtually any college. The Fifth edition provides an increased use of computing and graphical analysis throughout, without sacrificing concepts or rigor. In general, the 5e uses larger data sets in examples and exercises, and where methods can be automated within software without loss of understanding, it is so done.

Linear Regression Analysis

Application Guide

Distributions for Modeling Location, Scale, and Shape

Applied Linear Statistical Models

Generalized Linear Models for Insurance Rating

Import, Tidy, Transform, Visualize, and Model Data

This classic work by the late Stefan Banach has been translated into English so as to reach a yet wider audience. It contains the basics of the algebra of operators, concentrating on the study of linear operators, which corresponds to that of the linear forms a1x1 + a2x2 + ... + anxn of algebra. The book gathers results concerning linear operators defined in general spaces of a certain kind, principally in Banach spaces, examples of which are: the space of continuous functions, that of the pth-power-summable functions, Hilbert space, etc. The general theorems are interpreted in various mathematical areas, such as group theory, differential equations, integral equations, equations with infinitely many unknowns, functions of a real variable, summation methods and orthogonal series. A new fifty-page section (``Some Aspects of the Present Theory of Banach Spaces'') complements this important monograph.

In this monograph, authors Greg Taylor and Gráinne McGuire discuss generalized linear models (GLM) for loss reserving, beginning with strong emphasis on the chain ladder. The chain ladder is formulated in a GLM context, as is the statistical distribution of the loss reserve. This structure is then used to test the need for departure from the chain ladder model and to consider natural extensions of the chain ladder model that lend themselves to the GLM framework.

Many texts are excellent sources of knowledge about individual statistical tools, but the art of data analysis is about choosing and using multiple tools. Instead of presenting isolated techniques, this text emphasizes problem solving strategies that address the many issues arising when developing multivariable models using real data and not standard textbook examples. It includes imputation methods for dealing with missing data effectively, methods for dealing with nonlinear relationships and for making the estimation of transformations a formal part of the modeling process, methods for dealing with "too many variables to analyze and not enough observations," and powerful model validation techniques based on the bootstrap. This text realistically deals with model uncertainty and its effects on inference to achieve "safe data mining".

Non-life insurance pricing is the art of setting the price of an insurance policy, taking into consideration varoius properties of the insured object and the policy holder. Introduced by British actuaries generalized linear models (GLMs) have become today a the standard aproach for tariff analysis. The book focuses on methods based on GLMs that have been found useful in actuarial practice and provides a set of tools for a tariff analysis. Basic theory of GLMs in a tariff analysis setting is presented with useful extensions of standarde GLM theory that are not in common use. The book meets the European Core Syllabus for actuarial education and is written for actuarial students as well as practicing actuaries. To support reader real data of some complexity are provided at www.math.su.se/GLMbook.

With Applications to Linear Models, Logistic Regression, and Survival Analysis

An Introduction with R, Second Edition

Generalized Linear Models and Extensions, Second Edition

Vector Generalized Linear and Additive Models

Generalized Linear Models With Examples in R

With an Implementation in R

Wiley Series in Probability and Statistics A modern perspective on mixed models The availability of powerful computing methods in recent decades has thrust linear and nonlinear mixed models into the mainstream of statistical application. This volume offers a modern perspective on generalized, linear, and mixed models, presenting a unified and accessible treatment of the newest statistical methods for analyzing correlated, nonnormally distributed data. As a follow-up to Searle's classic, Linear Models, and Variance Components by Searle, Casella, and McCulloch, this new work progresses from the basic one-way classification to generalized linear mixed models. A variety of statistical methods are explained and illustrated, with an emphasis on maximum likelihood and restricted maximum likelihood. An invaluable resource for applied statisticians and industrial practitioners, as well as students interested in the latest results, Generalized, Linear, and Mixed Models features: * A review of the basics of linear models and linear mixed models * Descriptions of models for nonnormal data, including generalized linear and nonlinear models * Analysis and illustration of techniques for a variety of real data sets * Information on the accommodation of longitudinal data using these models * Coverage of the prediction of realized values of random effects * A discussion of the impact of computing issues on mixed models

Learn how to use R to turn raw data into insight, knowledge, and understanding. This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience, R for Data Science is designed to get you doing data science as quickly as possible. Authors Hadley Wickham and Garrett Grolemund guide you through the steps of importing, wrangling, exploring, and modeling your data and communicating the results. You'll get a complete, big-picture understanding of the data science cycle, along with basic tools you need to manage the details. Each section of the book is paired with exercises to help you practice what you've learned along the way. You'll learn how to: Wrangle—transform your datasets into a form convenient for analysis Program—learn powerful R tools for

solving data problems with greater clarity and ease Explore—examine your data, generate hypotheses, and quickly test them Model—provide a low-dimensional summary that captures true "signals" in your dataset Communicate—learn R Markdown for integrating prose, code, and results

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.

The author explains the theoretical underpinnings of generalized linear models so that researchers can decide how to select the best way to adapt their data for this type of analysis. Examples are provided to illustrate the application of GLM to actual data and the author includes his Web address where additional resources can be found.

Linear Models in Statistics

Theory and Computing

Nonparametric Regression and Generalized Linear Models

The Elements of Statistical Learning

Theory of Linear Operations

Dynamic Linear Models with R

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.

This book presents a statistical framework that expands generalized linear models (GLMs) for regression modelling. The framework shared in this book allows analyses based on many semi-traditional applied statistics models to be performed as a coherent whole. This is possible through the approximately half-a-dozen major classes of statistical models included in the book and the software infrastructure component, which makes the models easily operable. The book's methodology and accompanying software (the extensive VGAM R package) are directed at these limitations, and this is the first time the methodology and software are covered comprehensively in one volume. Since their advent in 1972, GLMs have unified important distributions under a single umbrella with enormous implications. The demands of practical data analysis, however, require a flexibility that GLMs do not have. Data-driven GLMs, in the form of generalized additive models (GAMs), are also largely confined to the exponential family. 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 may be used in senior undergraduate and first-year postgraduate courses on GLMs and regression modeling, including categorical data analysis. It may also serve as a reference on vector generalized linear models and as a methodology resource for VGAM users. The methodological contribution of this book stands alone and does not require use of the VGAM package. In the second part of the book, the R package VGAM makes applications of the methodology immediate. R code is integrated in the text, and datasets are used throughout. Potential applications include ecology, finance, biostatistics, and social sciences.

Build effective regression models in R to extract valuable insights from real data Key Features Implement different regression analysis techniques to solve common problems in data science - from data exploration to dealing with missing values From Simple Linear Regression to Logistic Regression - this book covers all regression techniques and their implementation in R A complete guide to building effective regression models in R and interpreting results from them to make valuable predictions Book Description Regression analysis is a statistical process which enables prediction of relationships between variables. The predictions are based on the casual effect of one variable upon another. Regression techniques for modeling and analyzing are employed on large set of data in order to reveal hidden relationship among the variables. This book will give you a rundown explaining what regression analysis is, explaining you the process from scratch. The first few chapters give an

understanding of what the different types of learning are – supervised and unsupervised, how these learnings differ from each other. We then move to covering the supervised learning in details covering the various aspects of regression analysis. The outline of chapters are arranged in a way that gives a feel of all the steps covered in a data science process – loading the training dataset, handling missing values, EDA on the dataset, transformations and feature engineering, model building, assessing the model fitting and performance, and finally making predictions on unseen datasets. Each chapter starts with explaining the theoretical concepts and once the reader gets comfortable with the theory, we move to the practical examples to support the understanding. The practical examples are illustrated using R code including the different packages in R such as R Stats, Caret and so on. Each chapter is a mix of theory and practical examples. By the end of this book you will know all the concepts and pain-points related to regression analysis, and you will be able to implement your learning in your projects. What you will learn Get started with the journey of data science using Simple linear regression Deal with interaction, collinearity and other problems using multiple linear regression Understand diagnostics and what to do if the assumptions fail with proper analysis Load your dataset, treat missing values, and plot relationships with exploratory data analysis Develop a perfect model keeping overfitting, under-fitting, and cross-validation into consideration Deal with classification problems by applying Logistic regression Explore other regression techniques – Decision trees, Bagging, and Boosting techniques Learn by getting it all in action with the help of a real world case study. Who this book is for This book is intended for budding data scientists and data analysts who want to implement regression analysis techniques using R. If you are interested in statistics, data science, machine learning and wants to get an easy introduction to the topic, then this book is what you need! Basic understanding of statistics and math will help you to get the most out of the book. Some programming experience with R will also be helpful

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

A Unified Approach

Introduction to Applied Linear Algebra

P-spline Vector Generalized Additive Models

Design and develop statistical nodes to identify unique relationships within data at scale

Generalized Linear Models for Insurance Data

The R Book

During the past decade there has been an explosion in computation and information technology. With it have come vast amounts of data in a variety of fields such as medicine, biology, finance, and marketing. The challenge of understanding these data has led to the development of new tools in the field of statistics, and spawned new areas such as data mining, machine learning, and bioinformatics. Many of these tools have common underpinnings but are often expressed with different terminology. This book describes the important ideas in these areas in a common conceptual framework. While the approach is statistical, the emphasis is on concepts rather than mathematics. Many examples are given, with a liberal use of color graphics. It should be a valuable resource for statisticians and anyone interested in data mining in science or industry. The book's coverage is broad, from supervised learning (prediction) to unsupervised learning. The many topics include neural networks, support vector machines, classification trees and boosting---the first comprehensive treatment of this topic in any book. This

major new edition features many topics not covered in the original, including graphical models, random forests, ensemble methods, least angle regression & path algorithms for the lasso, non-negative matrix factorization, and spectral clustering. There is also a chapter on methods for "wide" data (p bigger than n), including multiple testing and false discovery rates. Trevor Hastie, Robert Tibshirani, and Jerome Friedman are professors of statistics at Stanford University. They are prominent researchers in this area: Hastie and Tibshirani developed generalized additive models and wrote a popular book of that title. Hastie co-developed much of the statistical modeling software and environment in R/S-PLUS and invented principal curves and surfaces. Tibshirani proposed the lasso and is co-author of the very successful An Introduction to the Bootstrap. Friedman is the co-inventor of many data-mining tools including CART, MARS, projection pursuit and gradient boosting.

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.

Vector Generalized Linear and Additive ModelsWith an Implementation in RSpringer

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.

Non-Life Insurance Pricing with Generalized Linear Models

Statistical Regression and Classification

From Linear Models to Machine Learning

Applied Regression Analysis and Generalized Linear Models

Statistical Learning with Sparsity

Spatial Regression Analysis Using Eigenvector Spatial Filtering provides theoretical foundations and guides practical implementation of the Moran eigenvector spatial filtering (MESF) technique. MESF is a novel and powerful spatial statistical methodology that allows spatial scientists to account for spatial autocorrelation in their georeferenced data analyses. Its appeal is in its simplicity, yet its implementation drawbacks include serious complexities associated with constructing an eigenvector spatial filter. This book discusses MESF specifications for various intermediate-level topics, including spatially varying coefficients models, (non) linear mixed models, local spatial autocorrelation, space-time models, and spatial interaction models. Spatial Regression Analysis Using Eigenvector Spatial Filtering is accompanied by sample R codes and a Windows application with illustrative datasets so that readers can replicate the examples in the book and apply the methodology to their own application projects. It also includes a Foreword by Pierre Legendre. Reviews the uses of ESF across

linear regression, generalized linear regression, spatial autocorrelation measurement, and spatially varying coefficient models Includes computer code and template datasets for further modeling Provides comprehensive coverage of related concepts in spatial data analysis and spatial statistics 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.

"This volume presents in detail the fundamental theories of linear regression analysis and diagnosis, as well as the relevant statistical computing techniques so that readers are able to actually model the data using the techniques described in the book. This book is suitable for graduate students who are either majoring in statistics/biostatistics or using linear regression analysis substantially in their subject area." --Book Jacket.

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.
Data Mining, Inference, and Prediction
Nonparametric and Semiparametric Models
Regression Analysis with R
Regression Modeling Strategies
A roughness penalty approach
Vectors, Matrices, and Least Squares

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 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.

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.

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.

An Introduction with R

An Introduction to Statistical Learning

Applying Generalized Linear Models

Stochastic Loss Reserving Using Generalized Linear Models

Generalized Linear Mixed Models

Mixed Effects Models for Complex Data

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.

In recent years, there has been a great deal of interest and activity in the general area of nonparametric smoothing in statistics. This monograph concentrates on the roughness penalty method and shows how this technique provides a unifying approach to a wide range of smoothing problems. The method allows parametric assumptions to be realized in regression problems, in those approached by generalized linear modelling, and in many other contexts. The emphasis throughout is methodological rather than theoretical, and it concentrates on statistical and computation issues. Real data examples are used to illustrate the various methods and to compare them with standard parametric approaches. Some publicly available software is also discussed. The mathematical treatment is self-contained and depends mainly on simple linear algebra and calculus. This monograph will be useful both as a reference work for research and applied statisticians and as a text for graduate students and other encountering the material for the first time.

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.

The statistical and mathematical principles of smoothing with a focus on applicable techniques are presented in this book. It naturally splits into two parts: The first part is intended for undergraduate students majoring in mathematics, statistics, econometrics or biometrics whereas the second part is intended to be used by master and PhD students or researchers. The material is easy to accomplish since the e-book character of the text gives a maximum of flexibility in learning (and teaching) intensity.

XploRe® - Application Guide

Flexible Regression and Smoothing

Interpretable Machine Learning

with Applications in R

R for Data Science

Spatial Regression Analysis Using Eigenvector Spatial Filtering

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.

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. 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. 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 linear statistical models. The book presents a broad, in-depth overview of the most commonly used statistical models by discussing the theory underlying the models, R software applications, and examples with crafted models to elucidate key ideas and promote practical model building. 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 of Linear 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.

The Lasso and Generalizations

Using GAMLSS in R

Foundations of Linear and Generalized Linear Models

Generalized Additive Models

Generalized Linear Models

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.

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.