Generalized Linear Models For Insurance Data International Series On Actuarial Science

Concerned with the use of generalised linear models for univariate and multivariate regression analysis, this is a detailed introductory survey of the subject, based on the analysis of real data drawn from a variety of subjects such as the biological sciences, economics, and the social sciences. Where possible, technical details and proofs are deferred to an appendix in order to provide an accessible account for non-experts. Topics covered include: models for multi-categorical responses, model checking, time series and longitudinal data, random effects models, and state-space models. Throughout, the authors have taken great pains to discuss the underlying theoretical ideas in ways that relate well to the data at hand. As a result, numerous researchers whose work relies on the use of these models will find this an invaluable account.

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.

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 foundations and applications for quantitatively oriented graduate social-science students and individual researchers.

The success of the first edition of Generalized Linear Models led to the updated Second Edition, which continues to provide a definitive unified, treatment of methods for the analysis of diverse types of data. Today, it remains popular for its clarity, richness of content and direct relevance to agricultural, biological, health, engineering, and ot

Generalized Linear Models for Insurance DataCambridge University Press

Generalized linear models (GLMs) are gaining popularity as a statistical analysis method for insurance data. We study the theory and applications of GLMs in insurance. The first chapter gives an introduction of the theory of GLMs and generalized linear mixed models (GLMMs) as well as the bias correction for GLM estimators. It is shown that the maximum likelihood estimators (MLEs) of

the parameters in GLMs are asymptotically normal and asymptotically unbiased. However, when the sample size n or the total Fisher information is small, the MLEs can be biased. The bias is usually ignored in practice. However, in small or moderate--size portfolios, a bias correction can be appreciable. For segmented portfolios, as in car insurance, the question of credibility arises naturally; how many observations are needed in a risk class before the GLM estimators can be considered credible? In this thesis we study the limited fluctuations credibility of the GLM estimators as well as in the extended case of GLMMs. We show how credibility depends on the sample size, the distribution of covariates and the link function. We give a criteria for full credibility of the GLM estimators. This provides a mechanism to obtain confidence intervals for the GLM and GLMM estimators. If the full credibility criteria cannot be satisfied, it is interesting to calculate the partial credibility matrix and the GLM estimators. Here, for a general link function the credibility matrix is not known explicitly. Under certain assumptions, numerical methods are developed to compute the GLM estimators and the credibility matrix. For some specific link functions, further properties are developed. For instance, Hachemeister's credibility regression model is one such special case of our model, where the link function is linear. Loss

reserving is a major challenge for casualty actuaries due to the frequently changing business environments. Recently, some aggregate loss reserving models have been extended to or developed by research actuaries within the framework of GLMs. In this thesis we establish a structural loss reserving model which combines the exposure and loss emergence patterns and the loss development pattern, again within the framework of a GLM. Discounted loss reserves can also be obtained from this model.

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 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.

Modern Actuarial Risk Theory contains what every actuary needs to know about non-life insurance mathematics. It starts with the standard material like utility theory, individual and collective model and basic ruin theory. Other topics are risk measures and premium principles, bonus-malus systems, ordering of risks and credibility theory. It also contains some chapters about Generalized Linear Models, applied to rating and IBNR problems. As to the level of the mathematics, the book would fit in a bachelors or masters program in quantitative economics or mathematical statistics. This second and.

The book develops the capabilities arising from the cooperation between mathematicians and statisticians working in insurance and finance fields. It gathers some of the papers presented at the conference MAF2010, held in Ravello (Amalfi coast), and successively, after a reviewing process, worked out to this aim.

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.
Statistical and Probabilistic Methods in Actuarial Science covers many of the diverse methods in applied probability and statistics for students aspiring to careers in insurance, actuarial science, and finance. The book builds on students' existing knowledge of probability and statistics by establishing a solid and thorough understanding of
A Hands-On Way to Learning Data AnalysisPart of the core of statistics, linear models are used to make predictions and explain the relationship between the response and the predictors. Understanding linear models is crucial to a broader competence in the practice of statistics. Linear Models with R, Second Edition explains how to use linear models
A rigorous, self-contained examination of mixed model theory and application Mixed modeling is one of the most promising and exciting areas of statistical analysis, enabling the analysis of nontraditional, clustered data that may come in the form of shapes or images. This book provides in-depth mathematical coverage of mixed models' statistical properties and numerical algorithms, as well as applications such as the analysis of tumor regrowth, shape, and image. Paying special attention to algorithms and their implementations, the book discusses: Modeling of complex clustered or longitudinal data Modeling data with multiple sources of variation Modeling biological variety and heterogeneity Mixed model as a compromise between the frequentist and Bayesian approaches Mixed model for the penalized log-likelihood Healthy Akaike Information Criterion (HAIC) How to cope with parameter multidimensionality How to solve ill-posed

problems including image reconstruction problems Modeling of ensemble shapes and images Statistics of image processing Major results and points of discussion at the end of each chapter along with "Summary Points" sections make this reference not only comprehensive but also highly accessible for professionals and students alike in a broad range of fields such as cancer research, computer science, engineering, and industry.

Predictive modeling uses data to forecast future events. It exploits relationships between explanatory variables and the predicted variables from past occurrences to predict future outcomes. Forecasting financial events is a core skill that actuaries routinely apply in insurance and other risk-management applications. Predictive Modeling Applications in Actuarial Science emphasizes life-long learning by developing tools in an insurance context, providing the relevant actuarial applications, and introducing advanced statistical techniques that can be used to gain a competitive advantage in situations with complex data. Volume 2 examines applications of predictive modeling. Where Volume 1 developed the foundations of predictive modeling, Volume 2 explores practical uses for techniques, focusing on property and casualty insurance. Readers are exposed to a variety of techniques in concrete, real-life contexts that demonstrate their value and the overall value of predictive modeling, for seasoned practicing analysts as well as those just starting out.

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.

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.

A Hands-On Approach to Understanding and Using Actuarial Models Computational Actuarial Science with R provides an introduction to the computational aspects of actuarial science. Using simple R code, the book helps you understand the algorithms involved in actuarial computations. It also covers more advanced topics, such as parallel computing and C/C++ embedded codes. After an introduction to the R language, the book is divided into four parts. The first one addresses methodology and statistical modeling issues. The second part discusses the computational facets of life insurance, including life contingencies calculations and prospective life tables. Focusing on finance from an actuarial perspective, the next part presents techniques for modeling stock prices, nonlinear time series, yield curves, interest rates, and portfolio optimization. The last part explains how to use R to deal with computational issues of nonlife insurance. Taking a do-it-yourself approach to understanding algorithms, this book demystifies the computational aspects of actuarial science. It shows that even complex computations can usually be done without too much trouble. Datasets used in the text are available in an R package (CASdatasets).

Generalized Linear Models for Categorical and Continuous Limited Dependent Variables is designed for graduate students and researchers in the behavioral, social, health, and medical sciences. It incorporates examples of truncated counts, censored continuous variables, and doubly bounded continuous variables, such as percentages. The book provides broad, but unified, coverage, and the authors integrate the concepts and ideas shared across models and types of data, especially regarding conceptual links between discrete and continuous limited dependent variables. The authors argue that these dependent variables are, if anything, more common throughout the human sciences than the kind that suit linear regression. They cover special cases or extensions of models, estimation methods, model diagnostics, and, of course, software. They also discuss bounded continuous variables, boundary-inflated models, and methods for modeling heteroscedasticity. Wherever possible, the authors have illustrated concepts, models, and techniques with real or realistic datasets and demonstrations in R and Stata, and each chapter includes several exercises at the end. The illustrations and exercises help readers build conceptual understanding and fluency in using these techniques. At several points the authors bring together material that has been previously scattered across the literature in journal articles, software package documentation files, and blogs. These features help students learn to choose the appropriate models for their purpose.

This book is for actuaries and financial analysts developing their expertise in statistics and who wish to become familiar with concrete examples of predictive modeling.

This book summarizes the state of the art in generalized linear models (GLMs) and their various extensions: GAMs, mixed models and credibility, and

some nonlinear variants (GNMs). In order to deal with tail events, analytical tools from Extreme Value Theory are presented. Going beyond mean modeling, it considers volatility modeling (double GLMs) and the general modeling of location, scale and shape parameters (GAMLSS). Actuaries need these advanced analytical tools to turn the massive data sets now at their disposal into opportunities. The exposition alternates between methodological aspects and case studies, providing numerical illustrations using the R statistical software. The technical prerequisites are kept at a reasonable level in order to reach a broad readership. This is the first of three volumes entitled Effective Statistical Learning Methods for Actuaries. Written by actuaries for actuaries, this series offers a comprehensive overview of insurance data analytics with applications to P&C, life and health insurance. Although closely related to the other two volumes, this volume can be read independently.

The volume presents innovations in data analysis and classification and gives an overview of the state of the art in these scientific fields and applications. Areas that receive considerable attention in the book are discrimination and clustering, data analysis and statistics, as well as applications in marketing, finance, and medicine. The reader will find material on recent technical and methodological developments and a large number of applications

demonstrating the usefulness of the newly developed techniques.

A wide range of topics to give students a firm foundation in statistical and actuarial concepts and their applications.

This collection of articles addresses the most modern forms of loss reserving methodology: granular models and machine learning models. New methodologies come with questions about their applicability. These questions are discussed in one article, which focuses on the relative merits of granular and machine learning models. Others illustrate applications with real-world data. The examples include neural networks, which, though well known in some disciplines, have previously been limited in the actuarial literature. This volume expands on that literature, with specific attention to their application to loss reserving. For example, one of the articles introduces the application of neural networks of the gated recurrent unit form to the actuarial literature, whereas another uses a penalized neural network. Neural networks are not the only form of machine learning, and two other papers outline applications of gradient boosting and regression trees respectively. Both articles construct loss reserves at the individual claim level so that these models resemble granular models. One of these articles provides a practical application of the model to claim watching, the action of monitoring

claim development and anticipating major features. Such watching can be used as an early warning system or for other administrative purposes. Overall, this volume is an extremely useful addition to the libraries of those working at the loss reserving frontier.

This book teaches multiple regression and time series and how to use these to analyze real data in risk management and finance.

Finance and insurance companies are facing a wide range of parametric statistical problems. Statistical experiments generated by a sample of independent and identically distributed random variables are frequent and well understood, especially those consisting of probability measures of an exponential type. However, the aforementioned applications also offer non-classical experiments implying observation samples of independent but not identically distributed random variables or even dependent random variables. Three examples of such experiments are treated in this book. First, the Generalized Linear Models are studied. They extend the standard regression model to non-Gaussian distributions. Statistical experiments with Markov chains are considered next. Finally, various statistical experiments generated by fractional Gaussian noise are also described. In this book, asymptotic properties of several sequences of estimators are detailed. The notion of asymptotical

efficiency is discussed for the different statistical experiments considered in order to give the proper sense of estimation risk. Eighty examples and computations with R software are given throughout the text. Examines a range of statistical inference methods in the context of finance and insurance applications Presents the LAN (local asymptotic normality) property of likelihoods Combines the proofs of LAN property for different statistical experiments that appears in financial and insurance mathematics Provides the proper description of such statistical experiments and invites readers to seek optimal estimators (performed in R) for such statistical experiments

Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R is designed for undergraduate students who have successfully completed a multiple linear regression course, helping them develop an expanded modeling toolkit that includes non-normal responses and correlated structure. Even though there is no mathematical prerequisite, the authors still introduce fairly sophisticated topics such as likelihood theory, zero-inflated Poisson, and parametric bootstrapping in an intuitive and applied manner. The case studies and exercises feature real data and real research questions; thus, most of the data in the textbook comes from collaborative research conducted by the authors and their students, or from student projects.

Every chapter features a variety of conceptual exercises, guided exercises, and open-ended exercises using real data. After working through this material, students will develop an expanded toolkit and a greater appreciation for the wider world of data and statistical modeling. A solutions manual for all exercises is available to qualified instructors at the book's website at www.routledge.com, and data sets and Rmd files for all case studies and exercises are available at the authors' GitHub repo (https://github.com/proback/BeyondMLR)

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.

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.

The book contains important material on topics that are relevant for recent insurance and actuarial developments including determining solvency measures, fair-value computations, reserving, ranking of risks, modelling dependencies and the use of generalized linear models. Numerous exercises and the hints for solving them make the book useful as a textbook. Practical paradigms in insurance are presented in a way that is appealing to actuaries in their daily business.