

Data Analysis And Modeling Techniques

BIG DATA, ARTIFICIAL INTELLIGENCE AND DATA ANALYSIS SET Coordinated by Jacques Janssen Data analysis is a scientific field that continues to grow enormously, most notably over the last few decades, following rapid growth within the tech industry, as well as the wide applicability of computational techniques alongside new advances in analytic tools. Modeling enables data analysts to identify relationships, make predictions, and to understand, interpret and visualize the extracted information more strategically. This book includes the most recent advances on this topic, meeting increasing demand from wide circles of the scientific community. Applied Modeling Techniques and Data Analysis 2 is a collective work by a number of leading scientists, analysts, engineers, mathematicians and statisticians, working on the front end of data analysis and modeling applications. The chapters cover a cross section of current concerns and research interests in the above scientific areas. The collected material is divided into appropriate sections to provide the reader with both theoretical and applied information on data analysis methods, models and techniques, along with appropriate applications.

Software Systems Architecture is a practitioner-oriented guide to designing and implementing effective architectures for information systems. It is both a readily accessible introduction to software architecture and an invaluable handbook of well-established best practices. It shows why the role of the architect is central to any successful information-systems development project, and, by presenting a set of architectural viewpoints and perspectives, provides specific direction for improving your own and your organization's approach to software systems architecture. With this book you will learn how to Design an architecture that reflects and balances the different needs of its stakeholders Communicate the architecture to stakeholders and demonstrate that it has met their requirements Focus on architecturally significant aspects of design, including frequently overlooked areas such as performance, resilience, and location Use scenarios and patterns to drive the creation and validation of your architecture Document your architecture as a set of related views Use perspectives to ensure that your architecture exhibits important qualities such as performance, scalability, and security The architectural viewpoints and perspectives presented in the book also provide a valuable long-term reference source for new and experienced architects alike. Whether you are an aspiring or practicing software architect, you will find yourself referring repeatedly to the practical advice in this book throughout the lifecycle of your projects. A supporting Web site containing further information can be found at www.viewpoints-and-perspectives.info

An Applied Treatment of Modern Graphical Methods for Analyzing Categorical Data Discrete 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 text employs the latest ideas in teaching business statistics and follows the philosophy espoused at the conference "Making Statistics More Effective in Schools of Business" (MSMESB). It emphasizes modern statistical methods and data analysis with a decreased emphasis on classical hypothesis testing and probability. It presents a problem-solving approach to the analysis of real data sets and procedures for data collection, design, and interpretation. It covers statistics in the context of the scientific method for problem recognition, problem formulation, and problem solving. Concrete examples of statistical techniques and computer use give students a practical framework of business statistics in practice.

Multilevel modeling is an increasingly popular multivariate technique that is widely applied in the social sciences. Increasingly, practitioners are making instructional decisions based on results from their multivariate analyses, which often come from nested data that lend themselves to multilevel modeling techniques. As data-driven decision making becomes more critical to colleges and universities, multilevel modeling is a tool that will lead to more efficient estimates and enhance understanding of complex relationships. This volume illustrates both the theoretical underpinnings and practical applications of multilevel modeling in IR. It introduces the fundamental concepts of multilevel modeling techniques in a conceptual and technical manner. Providing a range of examples of nested models that are based on linear and categorical outcomes, it then offers important suggestions about presenting results of multilevel models through charts and graphs. This is the 154th volume of this Jossey-Bass quarterly report series. Always timely and comprehensive, New Directions for Institutional Research provides planners and administrators in all types of academic institutions with guidelines in such areas as resource coordination, information analysis, program evaluation, and institutional management.

Including new developments and publications which have appeared since the publication of the first edition in 1995, this second edition: *gives a comprehensive introductory account of event history modeling techniques and their use in applied research in economics and the social sciences; *demonstrates that event history modeling is a major step forward in causal analysis. To do so the authors show that event history models employ the time-path of changes in states and relate changes in causal variables in the past to changes in discrete outcomes in the future; and *introduces the reader to the computer program Transition Data Analysis (TDA). This software estimates the sort of models most frequently used with longitudinal data, in particular, discrete-time and continuous-time event history data. Techniques of Event History Modeling can serve as a student textbook in the fields of statistics, economics, the social sciences, psychology, and the political sciences. It can also be used as a reference for scientists in all fields of research.

Concise, complete, nontechnical; the ideal introduction to an increasingly important topic In recent years, the use of statistical methods for categorical data has increased dramatically in a variety of areas and applications. This book provides an applied introduction to the most important methods for analyzing categorical data. It summarizes methods that have long played a prominent role, such as chi-squared tests, but places special emphasis on logistic regression and loglinear modeling techniques. Special features of the book include: Emphasis on logistic regression modeling of binary data and Poisson regression modeling of count data A unified perspective, based on generalized linear models, that connects these methods with standard regression methods for normally-distributed data An appendix showing the use of a new SAS procedure (GENMOD) for generalized linear modeling that can conduct nearly all methods presented in the book An entertaining historical perspective of the development of the methods Specialized methods for ordinal data, small samples, multicategory data, and matched pairs More than 100 examples of real data sets and more than 200 exercises Writing in an applied, nontechnical style, Alan Agresti illustrates methods using a wide variety of real data, including alcohol, cigarette, and marijuana use by teenagers; AZT use and delay of AIDS; space shuttle launches and O-ring failure; passive smoking and lung cancer; and much more. An Introduction to Categorical Data Analysis is an invaluable tool for social, behavioral, and biomedical scientists, as well as researchers in public health, marketing, education, biological and agricultural sciences, and industrial quality control. An Instructor's Manual presenting detailed solutions to all the problems in the book is available from the Wiley editorial department.

Using Python and R, the author addresses multiple business challenge, including segmentation, brand positioning, product choice modeling, pricing research, finance, sports, text analytics, sentiment analysis and social network analysis, cross sectional data, time series, spatial and spatio-temporal data.

Multivariate Analysis in the Pharmaceutical Industry provides industry practitioners with guidance on multivariate data methods and their applications over the lifecycle of a pharmaceutical product, from process development, to routine manufacturing, focusing on the challenges specific to each step. It includes an overview of regulatory guidance specific to the use of these methods, along with perspectives on the applications of these methods that allow for testing, monitoring and controlling products and processes. The book seeks to put multivariate analysis into a pharmaceutical context for the benefit of pharmaceutical practitioners, potential

practitioners, managers and regulators. Users will find a resources that addresses an unmet need on how pharmaceutical industry professionals can extract value from data that is routinely collected on products and processes, especially as these techniques become more widely used, and ultimately, expected by regulators. Targets pharmaceutical industry practitioners and regulatory staff by addressing industry specific challenges Includes case studies from different pharmaceutical companies and across product lifecycle of to introduce readers to the breadth of applications Contains information on the current regulatory framework which will shape how multivariate analysis (MVA) is used in years to come

Going beyond performing simple analyses, researchers involved in the highly dynamic field of computational intelligent data analysis design algorithms that solve increasingly complex data problems in changing environments, including economic, environmental, and social data. Computational Intelligent Data Analysis for Sustainable Development present Interest in predictive analytics of big data has grown exponentially in the four years since the publication of Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, Second Edition. In the third edition of this bestseller, the author has completely revised, reorganized, and repositioned the original chapters and produced 13 new chapters of creative and useful machine-learning data mining techniques. In sum, the 43 chapters of simple yet insightful quantitative techniques make this book unique in the field of data mining literature. What is new in the Third Edition: The current chapters have been completely rewritten. The core content has been extended with strategies and methods for problems drawn from the top predictive analytics conference and statistical modeling workshops. Adds thirteen new chapters including coverage of data science and its rise, market share estimation, share of wallet modeling without survey data, latent market segmentation, statistical regression modeling that deals with incomplete data, decile analysis assessment in terms of the predictive power of the data, and a user-friendly version of text mining, not requiring an advanced background in natural language processing (NLP). Includes SAS subroutines which can be easily converted to other languages. As in the previous edition, this book offers detailed background, discussion, and illustration of specific methods for solving the most commonly experienced problems in predictive modeling and analysis of big data. The author addresses each methodology and assigns its application to a specific type of problem. To better ground readers, the book provides an in-depth discussion of the basic methodologies of predictive modeling and analysis. While this type of overview has been attempted before, this approach offers a truly nitty-gritty, step-by-step method that both tyros and experts in the field can enjoy playing with.

Manage and work with business data effectively by learning data modeling techniques and leveraging the latest features of Power BI Key Features Understand data modeling techniques to get the best out of data using Power BI Define the relationships between data to extract valuable insights Solve a wide variety of business challenges by building optimal data models Book Description Microsoft Power BI is one of the most popular business intelligence tools available on the market for desktop and the cloud. This book will be your guide to understanding the ins and outs of data modeling and how to create data models using Power BI confidently. You'll learn how to connect data from multiple sources, understand data, define and manage relationships between data, and shape data models. In this book, you'll explore how to use data modeling and navigation techniques to define relationships and create a data model before defining new metrics and performing custom calculations using modeling features. As you advance through the chapters, the book will demonstrate how to create full-fledged data models, enabling you to create efficient data models and simpler DAX code with new data modeling features. With the help of examples, you'll discover how you can solve business challenges by building optimal data models and changing your existing data models to meet evolving business requirements. Finally, you'll learn how to use some new and advanced modeling features to enhance your data models to carry out a wide variety of complex tasks. By the end of this Power BI book, you'll have gained the skills you need to structure data coming from multiple sources in different ways to create optimized data models that support reporting and data analytics. What you will learn Implement virtual tables and time intelligence functionalities in DAX to build a powerful model Identify Dimension and Fact tables and implement them in Power Query Editor Deal with advanced data preparation scenarios while building Star Schema Explore best practices for data preparation and data modeling Discover different hierarchies and their common pitfalls Understand complex data models and how to decrease the level of model complexity with different data modeling approaches Who this book is for This MS Power BI book is for BI users, data analysts, and analysis developers who want to become well-versed with data modeling techniques to make the most of Power BI. Basic knowledge of Power BI and Star Schema will help you to understand the concepts covered in this book.

Praise for the First Edition "This...novel and highly stimulating book, which emphasizesolving real problems...should be widely read. It will have a positive and lasting effect on the teaching of modeling andstatistics in general." - Short Book Reviews This new edition features developments and real-worldexamples that showcase essential empirical modelingtechniques Successful empirical model building is founded on therelationship between data and approximate representations of thereal systems that generated that data. As a result, it is essentialfor researchers who construct these models to possess the specialskills and techniques for producing results that are insightful, reliable, and useful. Empirical Model Building: Data, Models, and Reality, Second Edition presents a hands-on approach to thebasic principles of empirical model building through a shrewdmixture of differential equations, computer-intensive methods, anddata. The book outlines both classical and new approaches andincorporates numerous real-world statistical problems thatillustrate modeling approaches that are applicable to a broad rangeof audiences, including applied statisticians and practicingengineers and scientists. The book continues to review models of growth and decay, systemswhere competition and interaction add to the complexitiy of themodel while discussing both classical and non-classical dataanalysis methods. This Second Edition now features further coverageof momentum based investing practices and resampling techniques,showcasing their importance and expediency in the real world. Theauthor provides applications of empirical modeling, such ascomputer modeling of the AIDS epidemic to explain why North Americahas most of the AIDS cases in the First World and data-basedstrategies that allow individual investors to build their owninvestment portfolios. Throughout the book, computer-based analysisis emphasized and newly added and updated exercises allow readersto test their comprehension of the presented material. Empirical Model Building, Second Edition is a suitablebook for modeling courses at the upper-undergraduate andgraduate levels. It is also an excellent reference for appliedstatisticians and researchers who carry out quantitative modelingin their everyday work.

The massive amount of current data has led to many different forms of data analysis processes that aim to explore this data to uncover valuable insights such as trends, anomalies and patterns. These processes support decision makers in their analysis of varied and changing data ranging from financial transactions to customer interactions and social network postings. These data

analysis processes use a wide variety of methods, including machine learning, in several domains such as business, finance, health and smart cities. Several data analysis processes have been proposed by academia and industry, including CRISP-DM and SEMMA, to describe the phases that data analysis experts go through when solving their problems. Specifically, CRISP-DM has modeling as one of its phases, which involves selecting a modeling technique, generating a test design, building a model, and assessing the model. However, automating these data analysis modeling processes faces numerous challenges, from a software engineering perspective. First, software users expect increased flexibility from the software as to the possible variations in techniques, types of data, and parameter settings. The software is required to accommodate complex usage and deployment variations, which are difficult for non-experts. Second, variability in functionality or quality attributes increases the complexity of these systems and makes them harder to design and implement. There is a lack of a framework design that takes variability into account. Third, the lack of a more comprehensive analysis of variability makes it difficult to evaluate opportunities for automating data analysis modeling. This thesis proposes a variability-aware design approach to the data analysis modeling process. The approach involves: (i) the assessment of the variabilities inherent in CRISP-DM data analysis modeling and the provision of feature models that represent these variabilities; (ii) the definition of a preliminary framework design that captures the identified variabilities; and (iii) evaluation of the framework design in terms of possibilities of automation. Overall, this work presents, to the best of our knowledge, the first approach based on variability assessment to design data modeling process such as CRISP-DM. The approach advances the state of the art by offering a variability-aware design a solution that can enhance system flexibility and a novel software design framework to support data analysis modeling.

SQL kann Spaß machen! Es ist ein erhebendes Gefühl, eine verworrene Datenmanipulation oder einen komplizierten Report mit einer einzigen Anweisung zu bewältigen und so einen Haufen Arbeit vom Tisch zu bekommen. Einführung in SQL bietet einen frischen Blick auf die Sprache, deren Grundlagen jeder Entwickler beherrschen muss. Die aktualisierte 2. Auflage deckt die Versionen MySQL 6.0, Oracle 11g und Microsoft SQL Server 2008 ab. Außerdem enthält sie neue Kapitel zu Views und Metadaten. SQL-Basics - in null Komma nichts durchstarten: Mit diesem leicht verständlichen Tutorial können Sie SQL systematisch und gründlich lernen, ohne sich zu langweilen. Es führt Sie rasch durch die Basics der Sprache und vermittelt darüber hinaus eine Reihe von häufig genutzten fortgeschrittenen Features. Mehr aus SQL-Befehlen herausholen: Alan Beaulieu will mehr vermitteln als die simple Anwendung von SQL-Befehlen: Er legt Wert auf ein tiefes Verständnis der SQL-Features und behandelt daher auch den Umgang mit Mengen, Abfragen innerhalb von Abfragen oder die überaus nützlichen eingebauten Funktionen von SQL. Die MySQL-Beispieldatenbank: Es gibt zwar viele Datenbankprodukte auf dem Markt, aber welches wäre zum Erlernen von SQL besser geeignet als MySQL, das weit verbreitete relationale Datenbanksystem? Der Autor hilft Ihnen, eine MySQL-Datenbank anzulegen, und nutzt diese für die Beispiele in diesem Buch. Übungen mit Lösungen: Zu jedem Thema finden Sie im Buch gut durchdachte Übungen mit Lösungen. So ist sichergestellt, dass Sie schnell Erfolgserlebnisse haben und das Gelernte auch praktisch umsetzen können.

Master predictive analytics, from start to finish Start with strategy and management Master methods and build models Transform your models into highly-effective code—in both Python and R This one-of-a-kind book will help you use predictive analytics, Python, and R to solve real business problems and drive real competitive advantage. You'll master predictive analytics through realistic case studies, intuitive data visualizations, and up-to-date code for both Python and R—not complex math. Step by step, you'll walk through defining problems, identifying data, crafting and optimizing models, writing effective Python and R code, interpreting results, and more. Each chapter focuses on one of today's key applications for predictive analytics, delivering skills and knowledge to put models to work—and maximize their value. Thomas W. Miller, leader of Northwestern University's pioneering program in predictive analytics, addresses everything you need to succeed: strategy and management, methods and models, and technology and code. If you're new to predictive analytics, you'll gain a strong foundation for achieving accurate, actionable results. If you're already working in the field, you'll master powerful new skills. If you're familiar with either Python or R, you'll discover how these languages complement each other, enabling you to do even more. All data sets, extensive Python and R code, and additional examples available for download at <http://www.ftpress.com/miller/> Python and R offer immense power in predictive analytics, data science, and big data. This book will help you leverage that power to solve real business problems, and drive real competitive advantage. Thomas W. Miller's unique balanced approach combines business context and quantitative tools, illuminating each technique with carefully explained code for the latest versions of Python and R. If you're new to predictive analytics, Miller gives you a strong foundation for achieving accurate, actionable results. If you're already a modeler, programmer, or manager, you'll learn crucial skills you don't already have. Using Python and R, Miller addresses multiple business challenges, including segmentation, brand positioning, product choice modeling, pricing research, finance, sports, text analytics, sentiment analysis, and social network analysis. He illuminates the use of cross-sectional data, time series, spatial, and spatio-temporal data. You'll learn why each problem matters, what data are relevant, and how to explore the data you've identified. Miller guides you through conceptually modeling each data set with words and figures; and then modeling it again with realistic code that delivers actionable insights. You'll walk through model construction, explanatory variable subset selection, and validation, mastering best practices for improving out-of-sample predictive performance. Miller employs data visualization and statistical graphics to help you explore data, present models, and evaluate performance. Appendices include five complete case studies, and a detailed primer on modern data science methods. Use Python and R to gain powerful, actionable, profitable insights about: Advertising and promotion Consumer preference and choice Market baskets and related purchases Economic forecasting Operations management Unstructured text and language Customer sentiment Brand and price Sports team performance And much more This book constitutes the thoroughly refereed joint post-proceedings of four international workshops held in conjunction with the 21st International Conference on Conceptual Modeling, ER 2002, in Tampere, Finland in October 2002. The 38 revised full papers presented were carefully selected and improved during two rounds of reviewing and revision. The papers are organized in topical sections on management of time and changes in information systems; architectures, models, and tools for systems evolution; conceptual modeling approaches to mobile information systems development; quality of conceptual models; requirements and entity relationship models; class models and architectures; Web and interactive models; processes, models, and Web services; e-business methods and technologies; and success factors for conceptual modeling in e-business.

This book provides a broad overview of basic multilevel modeling issues and illustrates techniques building analyses around several organizational data sets. Although the focus is primarily on educational and organizational settings, the examples will help the reader discover other applications for these techniques. Two basic classes of multilevel models are developed: multilevel

regression models and multilevel models for covariance structures--are used to develop the rationale behind these models and provide an introduction to the design and analysis of research studies using two multilevel analytic techniques--hierarchical linear modeling and structural equation modeling.

The availability of spatial databases and widespread use of geographic information systems has stimulated increasing interest in the analysis and modelling of spatial data. Spatial data analysis focuses on detecting patterns, and on exploring and modelling relationships between them in order to understand the processes responsible for their emergence. In this way, the role of space is emphasised, and our understanding of the working and representation of space, spatial patterns, and processes is enhanced. In applied research, the recognition of the spatial dimension often yields different and more meaningful results and helps to avoid erroneous conclusions. This book aims to provide an introduction into spatial data analysis to graduates interested in applied statistical research. The text has been structured from a data-driven rather than a theory-based perspective, and focuses on those models, methods and techniques which are both accessible and of practical use for graduate students. Exploratory techniques as well as more formal model-based approaches are presented, and both area data and origin-destination flow data are considered. Incorporates mixed-effects modeling techniques for more powerful and efficient methods This book presents current and effective nonparametric regression techniques for longitudinal data analysis and systematically investigates the incorporation of mixed-effects modeling techniques into various nonparametric regression models. The authors emphasize modeling ideas and inference methodologies, although some theoretical results for the justification of the proposed methods are presented. With its logical structure and organization, beginning with basic principles, the text develops the foundation needed to master advanced principles and applications. Following a brief overview, data examples from biomedical research studies are presented and point to the need for nonparametric regression analysis approaches. Next, the authors review mixed-effects models and nonparametric regression models, which are the two key building blocks of the proposed modeling techniques. The core section of the book consists of four chapters dedicated to the major nonparametric regression methods: local polynomial, regression spline, smoothing spline, and penalized spline. The next two chapters extend these modeling techniques to semiparametric and time varying coefficient models for longitudinal data analysis. The final chapter examines discrete longitudinal data modeling and analysis. Each chapter concludes with a summary that highlights key points and also provides bibliographic notes that point to additional sources for further study. Examples of data analysis from biomedical research are used to illustrate the methodologies contained throughout the book. Technical proofs are presented in separate appendices. With its focus on solving problems, this is an excellent textbook for upper-level undergraduate and graduate courses in longitudinal data analysis. It is also recommended as a reference for biostatisticians and other theoretical and applied research statisticians with an interest in longitudinal data analysis. Not only do readers gain an understanding of the principles of various nonparametric regression methods, but they also gain a practical understanding of how to use the methods to tackle real-world problems.

Praise for the Second Edition: "The authors present an intuitive and easy-to-read book. ... accompanied by many examples, proposed exercises, good references, and comprehensive appendices that initiate the reader unfamiliar with MATLAB." —Adolfo Alvarez Pinto, *International Statistical Review* "Practitioners of EDA who use MATLAB will want a copy of this book. ... The authors have done a great service by bringing together so many EDA routines, but their main accomplishment in this dynamic text is providing the understanding and tools to do EDA. —David A Huckaby, *MAA Reviews* Exploratory Data Analysis (EDA) is an important part of the data analysis process. The methods presented in this text are ones that should be in the toolkit of every data scientist. As computational sophistication has increased and data sets have grown in size and complexity, EDA has become an even more important process for visualizing and summarizing data before making assumptions to generate hypotheses and models. Exploratory Data Analysis with MATLAB, Third Edition presents EDA methods from a computational perspective and uses numerous examples and applications to show how the methods are used in practice. The authors use MATLAB code, pseudo-code, and algorithm descriptions to illustrate the concepts. The MATLAB code for examples, data sets, and the EDA Toolbox are available for download on the book's website. New to the Third Edition Random projections and estimating local intrinsic dimensionality Deep learning autoencoders and stochastic neighbor embedding Minimum spanning tree and additional cluster validity indices Kernel density estimation Plots for visualizing data distributions, such as beanplots and violin plots A chapter on visualizing categorical data

This text focuses on the use of smoothing methods for developing and estimating differential equations following recent developments in functional data analysis and building on techniques described in Ramsay and Silverman (2005) *Functional Data Analysis*. The central concept of a dynamical system as a buffer that translates sudden changes in input into smooth controlled output responses has led to applications of previously analyzed data, opening up entirely new opportunities for dynamical systems. The technical level has been kept low so that those with little or no exposure to differential equations as modeling objects can be brought into this data analysis landscape. There are already many texts on the mathematical properties of ordinary differential equations, or dynamic models, and there is a large literature distributed over many fields on models for real world processes consisting of differential equations. However, a researcher interested in fitting such a model to data, or a statistician interested in the properties of differential equations estimated from data will find rather less to work with. This book fills that gap.

Now, a leader of Northwestern University's prestigious analytics program presents a fully-integrated treatment of both the business and academic elements of marketing applications in predictive analytics. Writing for both managers and students, Thomas W. Miller explains essential concepts, principles, and theory in the context of real-world applications. Building on Miller's pioneering program, *Marketing Data Science* thoroughly addresses segmentation, target marketing, brand and product positioning, new product development, choice modeling, recommender systems, pricing research, retail site selection, demand estimation, sales forecasting, customer retention, and lifetime value analysis. Starting where Miller's widely-praised *Modeling Techniques in Predictive Analytics* left off, he integrates crucial information and insights that were previously segregated in texts on web analytics, network science, information technology, and programming. Coverage includes: The role of analytics in delivering effective messages on the web Understanding the web by understanding its hidden structures Being recognized on the web – and watching your own competitors Visualizing networks and understanding communities within them Measuring sentiment and making recommendations Leveraging

key data science methods: databases/data preparation, classical/Bayesian statistics, regression/classification, machine learning, and text analytics Six complete case studies address exceptionally relevant issues such as: separating legitimate email from spam; identifying legally-relevant information for lawsuit discovery; gleaning insights from anonymous web surfing data, and more. This text's extensive set of web and network problems draw on rich public-domain data sources; many are accompanied by solutions in Python and/or R. Marketing Data Science will be an invaluable resource for all students, faculty, and professional marketers who want to use business analytics to improve marketing performance.

To succeed with predictive analytics, you must understand it on three levels: Strategy and management Methods and models Technology and code This up-to-the-minute reference thoroughly covers all three categories. Now fully updated, this uniquely accessible book will help you use predictive analytics to solve real business problems and drive real competitive advantage. If you're new to the discipline, it will give you the strong foundation you need to get accurate, actionable results. If you're already a modeler, programmer, or manager, it will teach you crucial skills you don't yet have. Unlike competitive books, this guide illuminates the discipline through realistic vignettes and intuitive data visualizations—not complex math. Thomas W. Miller, leader of Northwestern University's pioneering program in predictive analytics, guides you through defining problems, identifying data, crafting and optimizing models, writing effective R code, interpreting results, and more. Every chapter focuses on one of today's key applications for predictive analytics, delivering skills and knowledge to put models to work—and maximize their value. Reflecting extensive student and instructor feedback, this edition adds five classroom-tested case studies, updates all code for new versions of R, explains code behavior more clearly and completely, and covers modern data science methods even more effectively. All data sets, extensive R code, and additional examples available for download at <http://www.ftpress.com/miller> If you want to make the most of predictive analytics, data science, and big data, this is the book for you. Thomas W. Miller's unique balanced approach combines business context and quantitative tools, appealing to managers, analysts, programmers, and students alike. Miller addresses multiple business cases and challenges, including segmentation, brand positioning, product choice modeling, pricing research, finance, sports, text analytics, sentiment analysis, and social network analysis. He illuminates the use of cross-sectional data, time series, spatial, and spatio-temporal data. You'll learn why each problem matters, what data are relevant, and how to explore the data you've identified. Miller guides you through conceptually modeling each data set with words and figures; and then modeling it again with realistic R programs that deliver actionable insights. You'll walk through model construction, explanatory variable subset selection, and validation, mastering best practices for improving out-of-sample predictive performance. Throughout, Miller employs data visualization and statistical graphics to help you explore data, present models, and evaluate performance. This edition adds five new case studies, updates all code for the newest versions of R, adds more commenting to clarify how the code works, and offers a more detailed and up-to-date primer on data science methods. Gain powerful, actionable, profitable insights about: Advertising and promotion Consumer preference and choice Market baskets and related purchases Economic forecasting Operations management Unstructured text and language Customer sentiment Brand and price Sports team performance And much more

This book offers a comprehensive and readable introduction to modern business and data analytics. It is based on the use of Excel, a tool that virtually all students and professionals have access to. The explanations are focused on understanding the techniques and their proper application, and are supplemented by a wealth of in-chapter and end-of-chapter exercises. In addition to the general statistical methods, the book also includes Monte Carlo simulation and optimization. The second edition has been thoroughly revised: new topics, exercises and examples have been added, and the readability has been further improved. The book is primarily intended for students in business, economics and government, as well as professionals, who need a more rigorous introduction to business and data analytics – yet also need to learn the topic quickly and without overly academic explanations.

Multilevel modelling is a data analysis method that is frequently used to investigate hierarchal data structures in educational, behavioural, health, and social sciences disciplines. Multilevel data analysis exploits data structures that cannot be adequately investigated using single-level analytic methods such as multiple regression, path analysis, and structural modelling. This text offers a comprehensive treatment of multilevel models for univariate and multivariate outcomes. It explores their similarities and differences and demonstrates why one model may be more appropriate than another, given the research objectives. New to this edition: An expanded focus on the nature of different types of multilevel data structures (e.g., cross-sectional, longitudinal, cross-classified, etc.) for addressing specific research goals; Varied modelling methods for examining longitudinal data including random-effect and fixed-effect approaches; Expanded coverage illustrating different model-building sequences and how to use results to identify possible model improvements; An expanded set of applied examples used throughout the text; Use of four different software packages (i.e., Mplus, R, SPSS, Stata), with selected examples of model-building input files included in the chapter appendices and a more complete set of files available online. This is an ideal text for graduate courses on multilevel, longitudinal, latent variable modelling, multivariate statistics, or advanced quantitative techniques taught in psychology, business, education, health, and sociology. Recommended prerequisites are introductory univariate and multivariate statistics.

The complexity, diversity, and random nature of transportation problems necessitates a broad analytical toolbox. Describing tools commonly used in the field, Statistical and Econometric Methods for Transportation Data Analysis, Second Edition provides an understanding of a broad range of analytical tools required to solve transportation problems. It includes a wide breadth of examples and case studies covering applications in various aspects of transportation planning, engineering, safety, and economics. After a solid refresher on statistical fundamentals, the book focuses on continuous dependent variable models and count and discrete dependent variable models. Along with an entirely new

section on other statistical methods, this edition offers a wealth of new material. New to the Second Edition A subsection on Tobit and censored regressions An explicit treatment of frequency domain time series analysis, including Fourier and wavelets analysis methods New chapter that presents logistic regression commonly used to model binary outcomes New chapter on ordered probability models New chapters on random-parameter models and Bayesian statistical modeling New examples and data sets Each chapter clearly presents fundamental concepts and principles and includes numerous references for those seeking additional technical details and applications. To reinforce a practical understanding of the modeling techniques, the data sets used in the text are offered on the book's CRC Press web page. PowerPoint and Word presentations for each chapter are also available for download.

This book inclusively and systematically presents the fundamental methods, models and techniques of practical application of grey data analysis, bringing together the authors' many years of theoretical exploration, real-life application, and teaching. It also reflects the majority of recent theoretical and applied advances in the theory achieved by scholars from across the world, providing readers a vivid overall picture of this new theory and its pioneering research activities. The book includes 12 chapters, covering the introduction to grey systems, a novel framework of grey system theory, grey numbers and their operations, sequence operators and grey data mining, grey incidence analysis models, grey clustering evaluation models, series of GM models, combined grey models, techniques for grey systems forecasting, grey models for decision-making, techniques for grey control, etc. It also includes a software package that allows practitioners to conveniently and practically employ the theory and methods presented in this book. All methods and models presented here were chosen for their practical applicability and have been widely employed in various research works. I still remember 1983, when I first participated in a course on Grey System Theory. The mimeographed teaching materials had a blue cover and were presented as a book. It was like finding a treasure: This fascinating book really inspired me as a young intellectual going through a period of confusion and lack of academic direction. It shone with pearls of wisdom and offered a beacon in the mist for a man trying to find his way in academic research. This book became the guiding light in my life journey, inspiring me to forge an indissoluble bond with Grey System Theory. —Sifeng Liu

Predictive modeling involves the use of data to forecast future events. It relies on capturing relationships between explanatory variables and the predicted variables from past occurrences and exploiting this to predict future outcomes. Forecasting future financial events is a core actuarial skill - actuaries routinely apply predictive-modeling techniques in insurance and other risk-management applications. This book is for actuaries and other financial analysts who are developing their expertise in statistics and wish to become familiar with concrete examples of predictive modeling. The book also addresses the needs of more seasoned practising analysts who would like an overview of advanced statistical topics that are particularly relevant in actuarial practice. Predictive Modeling Applications in Actuarial Science emphasizes lifelong learning by developing tools in an insurance context, providing the relevant actuarial applications, and introducing advanced statistical techniques that can be used by analysts to gain a competitive advantage in situations with complex data.

In recent years, spatio-temporal data has received a lot of attention and increasingly plays an important role in our everyday lives as we can witness from the fast-growing mobile technologies and its location-based application development. By spatio-temporal data, we mean data that is associated with specific spatial locations that change over time. For example, a cellphone or car with GPS will generate the object location at regular time intervals. Another example would be the track of a storm center as it moves. Spatio-temporal data could be thought of as a huge data warehouse, which contains hidden and meaningful information. However, to analyze the available spatiotemporal data directly from its original formats and locations is not easy because the data is often in a format that is difficult to analyze and is usually 'big'. Our research goals focus on spatio-temporal datasets and how to summarize, model, and conceptualize them for analysis and mining. Five main parts of this dissertation include: 1) spatio-temporal knowledge representation, 2) identifying meaningful concepts from raw data, 3) converting raw data to conceptual data, 4) analysis and mining of conceptual data, and 5) a general framework for big data analysis and mining. In the first part of the dissertation, we look at the spatio-temporal datasets in general by considering spatio-temporal data semantics using techniques similar to those utilized in the "Semantic Web". We work towards creating a spatio-temporal ontology framework, which can be used to represent and reason about spatio-temporal data. In the next three parts, we focus on the spatio-temporal datasets in a specific domain, which is rainfall precipitation data in the hydrology domain. However, the techniques and methodology that we use can be adapted to different types of hydrological data such as soil moisture, water level, etc., as well as other types of big spatio-temporal data. Therefore, in the final part, we propose a generalized framework for analyzing and mining big data in any given domain. The framework allows big data in a particular domain to be conceptually analyzed and mined by using ontologies and EER.

This book provides new theories, applications and quantitative methods in demography, population studies and statistics. It presents and applies data analysis, statistics and stochastic modeling techniques focusing on demography, population aging, mortality and health sciences. The book describes diverse stochastic processes as well as Markov and semi-Markov models in demography and population studies, along with chapters on statistical models and methods in biostatistics and epidemiology. As such the book will be a valuable source to demographers, health scientists, statisticians, economists and sociologists.

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. In fact, For most businesses and government agencies, lack of data isn't a problem. In fact, it's the opposite: there's often too much information available to make a clear decision. With so much data to sort through, you need something more from your data: the right data, accurate conclusions, data that informs your decision-making process. Being overwhelming is common but this book will help you to become better at data analysis with smart, high-tech methods and tips. The contents of this book are designed to help you use data to your advantage to enhance business outcomes! Here's what this book will teach you: Why data is your single most powerful tool How to conduct data analysis to enhance your business Which steps to take in performing predictive analysis What techniques you need to employ to achieve sustainable success PLUS Regression techniques Machine learning strategies Risk management tips And much, much, more To learn what you need to know; SCROLL UP AND DOWNLOAD YOUR COPY TODAY!

Applied Modeling Techniques and Data Analysis 1 Computational Data Analysis Methods and Tools John Wiley & Sons

BIG DATA, ARTIFICIAL INTELLIGENCE AND DATA ANALYSIS SET Coordinated by Jacques Janssen Data analysis is a scientific field that continues to grow enormously, most notably over the last few decades, following rapid growth within the tech industry, as well as the wide applicability of computational techniques alongside new advances in analytic tools. Modeling enables data analysts to identify relationships, make predictions, and to understand, interpret and visualize the extracted information more strategically. This book includes the most recent advances on this topic, meeting increasing demand from wide circles of the scientific community. Applied Modeling Techniques and Data Analysis 1 is a collective work by a number of leading scientists, analysts, engineers, mathematicians and statisticians, working on the front end of data analysis and modeling applications. The chapters cover a cross section of current concerns and research interests in the above scientific areas. The collected material is divided into appropriate sections to provide the reader with both theoretical and applied information on data analysis methods, models and techniques, along with appropriate applications.

Analytics and artificial intelligence (AI), what are they good for? The bandwagon keeps answering, absolutely everything! Analytics and artificial intelligence have captured the attention of everyone from top executives to the person in the street. While these disciplines have a relatively long history, within the last ten or so years they have exploded into corporate business and public consciousness. Organizations have rushed to embrace data-driven decision making. Companies everywhere are turning out products boasting that "artificial intelligence is included." We are indeed living in exciting times. The question we need to ask is, do we really know how to get business value from these exciting tools? Unfortunately, both the analytics and AI communities have not done a great job in collaborating and communicating with each other to build the necessary synergies. This book bridges the gap between these two critical fields. The book begins by explaining the commonalities and differences in the fields of data science, artificial intelligence, and autonomy by giving a historical perspective for each of these fields, followed by exploration of common technologies and current trends in each field. The book also readers introduces to applications of deep learning in industry with an overview of deep learning and its key architectures, as well as a survey and discussion of the main applications of deep learning. The book also presents case studies to illustrate applications of AI and analytics. These include a case study from the healthcare industry and an investigation of a digital transformation enabled by AI and analytics transforming a product-oriented company into one delivering solutions and services. The book concludes with a proposed AI-informed data analytics life cycle to be applied to unstructured data.

[Copyright: 84b51c96ddc79cd9188f62adfc3cd5a5](#)