Principal Components Analysis In R

Hands-On Machine Learning with RPractical Guide to Cluster Analysis in RMetabolomicsR for Data Science CookbookComputational Genomics With RHigh-Dimensional ProbabilityGeneralized Principal Component AnalysisData Analysis and Graphics Using RPrincipal Manifolds for Data Visualization and Dimension ReductionHands-On Programming with RMachine Learning EssentialsThree-mode Principal Component AnalysisCombining Soft Computing and Statistical Methods in Data AnalysisMultivariate Nonparametric Methods with RChemometrics with RPractical Guide To Principal Component Methods in RMethods of Multivariate AnalysisPrincipal Component AnalysisMastering Text Mining with RPrincipal Components AnalysisPractical Statistics for Data ScientistsPlaces Rated AlmanacIntroduction to Multivariate Statistical Analysis in ChemometricsBiplots in PracticeAn Introduction to Applied Multivariate Analysis with RExploratory Multivariate Analysis by Example Using Rggplot2Quantitative Investment Portfolio Analytics in RStatistics for Marketing and Consumer ResearchQuantitative Methods in Archaeology Using RModern Factor AnalysisAn R and S-Plus® Companion to Multivariate AnalysisStatisticsStatistics and Data Analysis in GeologyMultiple Factor Analysis by Example Using RModern Statistics for Modern BiologyMathematical and Statistical Methods in Food Science and TechnologyDeep Learning with RRobust StatisticsPrincipal Components Analysis

Hands-On Machine Learning with R

Since data grows faster than ever, the role of statistics becomes more and more crucial nowadays, and there is no doubt that statistics will be even more critical in the future. The application of statistics is extensive, and in our daily lives there is almost no human activity where the use of statistics is not needed. In this limited volume, we try to cover as many as different and multidisciplinary fields in statistics as possible and aim to present recent developments and applications of statistical analysis. Therefore, this book is organized into three sections: "The Role of Statistics on Quantification," "Applications of Statistics on Economics and Development," and "Applications of Statistics on Various Topics."

Practical Guide to Cluster Analysis in R

Metabolomics: Fundamentals and Applications authoritatively presents the basic principles and applications of metabolomics. Topics covered in this book range from the analysis of metabolites from different biological sources and their data processing and statistical analysis. This book serves as a basic guide for a wide range of audiences from less familiar with metabolomics techniques to more experienced researchers seeking to understand complex biological systems from the systems biology approach.

Metabolomics

Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment

system, write your own functions, and use all of R's programming tools. RStudio Master Instructor Garrett Grolemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time. Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users Use R programming tools such as if else statements, for loops, and S3 classes Learn how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them

R for Data Science Cookbook

Principal component analysis is probably the oldest and best known of the It was first introduced by Pearson (1901), techniques ofmultivariate analysis. and developed independently by Hotelling (1933). Like many multivariate methods, it was not widely used until the advent of electronic computers, but it is now well entrenched in virtually every statistical computer package. The central idea of principal component analysis is to reduce the dimen sionality of a data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. Computation of the principal components reduces to the solution of an eigenvalue-eigenvector problem for a positive-semidefinite symmetrie matrix. Thus, the definition and computation of principal components are straightforward but, as will be seen, this apparently simple technique has a wide variety of different applications, as well as a number of different deri vations. Any feelings that principal component analysis is a narrow subject should soon be dispelled by the present book; indeed some quite broad topics which are related to principal component analysis receive no more than a brief mention in the final two chapters.

Computational Genomics With R

High-Dimensional Probability

Statistical methods are a key part of of data science, yet very few data scientists have any formal statistics training. Courses and books on basic statistics rarely cover the topic from a data science perspective. This practical guide explains how to apply various statistical methods to data science, tells you how to avoid their misuse, and gives you advice on what's important and what's not. Many data science resources incorporate statistical methods but lack a deeper statistical perspective. If you're familiar with the R programming language, and have some exposure to statistics, this quick reference bridges the gap in an accessible, readable format. With this book, you'll learn: Why exploratory data analysis is a key preliminary step in data science How random sampling can reduce bias and

yield a higher quality dataset, even with big data How the principles of experimental design yield definitive answers to questions How to use regression to estimate outcomes and detect anomalies Key classification techniques for predicting which categories a record belongs to Statistical machine learning methods that "learn" from data Unsupervised learning methods for extracting meaning from unlabeled data

Generalized Principal Component Analysis

The first step-by-step guide to the quantitative analysis of archaeological data using the R statistical computing system.

Data Analysis and Graphics Using R

Amstat News asked three review editors to rate their topfive favorite books in the September 2003 issue. Methods ofMultivariate Analysis was among those chosen. When measuring several variables on a complex experimental unit,it is often necessary to analyze the variables simultaneously,rather than isolate them and consider them individually.Multivariate analysis enables researchers to explore the jointperformance of such variables and to determine the effect of eachvariable in the presence of the others. The Second Edition of AlvinRencher's Methods of Multivariate Analysis provides studentsof all statistical backgrounds with both the fundamental and moresophisticated skills necessary to master the discipline. To illustrate multivariate applications, the author providesexamples and exercises based on fifty-nine real data sets from awide variety of scientific fields. Rencher takes a "methods"approach to his subject, with an emphasis on how students andpractitioners can employ multivariate analysis in real-lifesituations. The Second Edition contains revised and updatedchapters from the critically acclaimed First Edition as well asbrand-new chapters on: Cluster analysis Multidimensional scaling Correspondence analysis Biplots Each chapter contains exercises, with corresponding answers andhints in the appendix, providing students the opportunity to testand extend their understanding of the subject. Methods ofMultivariate Analysis provides an authoritative reference forstatistics students as well as for practicing scientists andclinicians.

Principal Manifolds for Data Visualization and Dimension Reduction

This book provides a comprehensive introduction to the latest advances in the mathematical theory and computational tools for modeling high-dimensional data drawn from one or multiple low-dimensional subspaces (or manifolds) and potentially corrupted by noise, gross errors, or outliers. This challenging task requires the development of new algebraic, geometric, statistical, and computational methods for efficient and robust estimation and segmentation of one or multiple subspaces. The book also presents interesting real-world applications of these new methods in image processing, image and video segmentation, face recognition and clustering, and hybrid system identification etc. This book is intended to serve as a textbook for graduate students and beginning researchers in data science, machine learning, computer vision, image

and signal processing, and systems theory. It contains ample illustrations, examples, and exercises and is made largely self-contained with three Appendices which survey basic concepts and principles from statistics, optimization, and algebraic-geometry used in this book. René Vidal is a Professor of Biomedical Engineering and Director of the Vision Dynamics and Learning Lab at The Johns Hopkins University. Yi Ma is Executive Dean and Professor at the School of Information Science and Technology at ShanghaiTech University. S. Shankar Sastry is Dean of the College of Engineering, Professor of Electrical Engineering and Computer Science and Professor of Bioengineering at the University of California, Berkeley.

Hands-On Programming with R

Hands-on Machine Learning with R provides a practical and applied approach to learning and developing intuition into today's most popular machine learning methods. This book serves as a practitioner's guide to the machine learning process and is meant to help the reader learn to apply the machine learning stack within R, which includes using various R packages such as glmnet, h2o, ranger, xgboost, keras, and others to effectively model and gain insight from their data. The book favors a hands-on approach, providing an intuitive understanding of machine learning concepts through concrete examples and just a little bit of theory. Throughout this book, the reader will be exposed to the entire machine learning process including feature engineering, resampling, hyperparameter tuning, model evaluation, and interpretation. The reader will be exposed to powerful algorithms such as regularized regression, random forests, gradient boosting machines, deep learning, generalized low rank models, and more! By favoring a hands-on approach and using real word data, the reader will gain an intuitive understanding of the architectures and engines that drive these algorithms and packages, understand when and how to tune the various hyperparameters, and be able to interpret model results. By the end of this book, the reader should have a firm grasp of R's machine learning stack and be able to implement a systematic approach for producing high quality modeling results. Features: · Offers a practical and applied introduction to the most popular machine learning methods. · Topics covered include feature engineering, resampling, deep learning and more. · Uses a hands-on approach and real world data.

Machine Learning Essentials

For anyone in need of a concise, introductory guide to principal components analysis, this book is a must. Through an effective use of simple mathematical-geometrical and multiple real-life examples (such as crime statistics, indicators of drug abuse, and educational expenditures) -- and by minimizing the use of matrix algebra -- the reader can quickly master and put this technique to immediate use.

Three-mode Principal Component Analysis

Combining Soft Computing and Statistical Methods in Data Analysis

Multivariate Nonparametric Methods with R

Although there are several good books on principal component methods (PCMs) and related topics, we felt that many of them are either too theoretical or too advanced. This book provides a solid practical guidance to summarize, visualize and interpret the most important information in a large multivariate data sets, using principal component methods in R. The visualization is based on the factoextra R package that we developed for creating easily beautiful ggplot2-based graphs from the output of PCMs. This book contains 4 parts. Part I provides a quick introduction to R and presents the key features of FactoMineR and factoextra. Part II describes classical principal component methods to analyze data sets containing, predominantly, either continuous or categorical variables. These methods include: Principal Component Analysis (PCA, for continuous variables), simple correspondence analysis (CA, for large contingency tables formed by two categorical variables) and Multiple CA (MCA, for a data set with more than 2 categorical variables). In Part III, you'll learn advanced methods for analyzing a data set containing a mix of variables (continuous and categorical) structured or not into groups: Factor Analysis of Mixed Data (FAMD) and Multiple Factor Analysis (MFA). Part IV covers hierarchical clustering on principal components (HCPC), which is useful for performing clustering with a data set containing only categorical variables or with a mixed data of categorical and continuous variables.

Chemometrics with R

The book starts with the quote of the classical Pearson definition of PCA and includes reviews of various methods: NLPCA, ICA, MDS, embedding and clustering algorithms, principal manifolds and SOM. New approaches to NLPCA, principal manifolds, branching principal components and topology preserving mappings are described. Presentation of algorithms is supplemented by case studies. The volume ends with a tutorial PCA deciphers genome.

Practical Guide To Principal Component Methods in R

For anyone in need of a concise, introductory guide to principal components analysis, this book is a must. Through an effective use of simple mathematical-geometrical and multiple real-life examples (such as crime statistics, indicators of drug abuse, and educational expenditures) -- and by minimizing the use of matrix algebra -- the reader can quickly master and put this technique to immediate use.

Methods of Multivariate Analysis

Computational Genomics with R provides a starting point for beginners of genomic data analysis and also guides the more advanced practitioners to sophisticated data analysis techniques in genomics. The book covers topics from R programming to machine learning and statistics to the latest genomic data analysis techniques. The text provides accessible information and explanations always with the genomics context in the background. It also contains practical and well-documented examples in R so readers can analyze their data by simply reusing the

code presented. As the field of computational genomics is interdisciplinary, it requires different starting points for people with different backgrounds. A biologist might skip sections on basic genome biology and start with R programming whereas a computer scientist might want to start with genome biology. After reading, you will: - get the basics of R and dive right into specialized uses of R for computational genomics such as using Bioconductor packages. - be familiar with statistics, supervised and unsupervised learning techniques which are important in data modeling and exploratory analysis of high-dimensional data - understand genomic intervals and operations on them, which are used for tasks such as aligned read counting and genomic feature annotation - learn about the basics of processing and quality checking high-throughput sequencing data - be able to do sequence analysis, such as calculating GC content for parts of a genome or find transcription factor binding sites - learn about visualization techniques used in genomics, such as heatmaps, meta-gene plots, and genomic track visualization - be familiar with analysis of different high-throughput sequencing data sets, such as RNA-seq, ChIP-seq, and BS-seq - learn basic techniques to integrate and interpret multi-omics datasets

Principal Component Analysis

Introduces deep learning systems using the powerful Keras library and its R language interface. The book builds your understanding of deep learning through intuitive explanations and practical examples.

Mastering Text Mining with R

Multiple factor analysis (MFA) enables users to analyze tables of individuals and variables in which the variables are structured into quantitative, qualitative, or mixed groups. Written by the co-developer of this methodology, Multiple Factor Analysis by Example Using R brings together the theoretical and methodological aspects of MFA. It also includes examples of applications and details of how to implement MFA using an R package (FactoMineR). The first two chapters cover the basic factorial analysis methods of principal component analysis (PCA) and multiple correspondence analysis (MCA). The next chapter discusses factor analysis for mixed data (FAMD), a little-known method for simultaneously analyzing quantitative and qualitative variables without group distinction. Focusing on MFA, subsequent chapters examine the key points of MFA in the context of quantitative variables as well as qualitative and mixed data. The author also compares MFA and Procrustes analysis and presents a natural extension of MFA: hierarchical MFA (HMFA). The final chapter explores several elements of matrix calculation and metric spaces used in the book.

Principal Components Analysis

High-dimensional probability offers insight into the behavior of random vectors, random matrices, random subspaces, and objects used to quantify uncertainty in high dimensions. Drawing on ideas from probability, analysis, and geometry, it lends itself to applications in mathematics, statistics, theoretical computer science, signal processing, optimization, and more. It is the first to integrate theory, key

tools, and modern applications of high-dimensional probability. Concentration inequalities form the core, and it covers both classical results such as Hoeffding's and Chernoff's inequalities and modern developments such as the matrix Bernstein's inequality. It then introduces the powerful methods based on stochastic processes, including such tools as Slepian's, Sudakov's, and Dudley's inequalities, as well as generic chaining and bounds based on VC dimension. A broad range of illustrations is embedded throughout, including classical and modern results for covariance estimation, clustering, networks, semidefinite programming, coding, dimension reduction, matrix completion, machine learning, compressed sensing, and sparse regression.

Practical Statistics for Data Scientists

Using formal descriptions, graphical illustrations, practical examples, and R software tools, Introduction to Multivariate Statistical Analysis in Chemometrics presents simple yet thorough explanations of the most important multivariate statistical methods for analyzing chemical data. It includes discussions of various statistical methods, such as principal component analysis, regression analysis, classification methods, and clustering. Written by a chemometrician and a statistician, the book reflects the practical approach of chemometrics and the more formally oriented one of statistics. To enable a better understanding of the statistical methods, the authors apply them to real data examples from chemistry. They also examine results of the different methods, comparing traditional approaches with their robust counterparts. In addition, the authors use the freely available R package to implement methods, encouraging readers to go through the examples and adapt the procedures to their own problems. Focusing on the practicality of the methods and the validity of the results, this book offers concise mathematical descriptions of many multivariate methods and employs graphical schemes to visualize key concepts. It effectively imparts a basic understanding of how to apply statistical methods to multivariate scientific data.

Places Rated Almanac

The majority of data sets collected by researchers in all disciplines are multivariate, meaning that several measurements, observations, or recordings are taken on each of the units in the data set. These units might be human subjects, archaeological artifacts, countries, or a vast variety of other things. In a few cases, it may be sensible to isolate each variable and study it separately, but in most instances all the variables need to be examined simultaneously in order to fully grasp the structure and key features of the data. For this purpose, one or another method of multivariate analysis might be helpful, and it is with such methods that this book is largely concerned. Multivariate analysis includes methods both for describing and exploring such data and for making formal inferences about them. The aim of all the techniques is, in general sense, to display or extract the signal in the data in the presence of noise and to find out what the data show us in the midst of their apparent chaos. An Introduction to Applied Multivariate Analysis with R explores the correct application of these methods so as to extract as much information as possible from the data at hand, particularly as some type of graphical representation, via the R software. Throughout the book, the authors give many examples of R code used to apply the multivariate techniques to

multivariate data.

Introduction to Multivariate Statistical Analysis in Chemometrics

Master text-taming techniques and build effective text-processing applications with R About This Book Develop all the relevant skills for building text-mining apps with R with this easy-to-follow guide Gain in-depth understanding of the text mining process with lucid implementation in the R language Example-rich guide that lets you gain high-quality information from text data Who This Book Is For If you are an R programmer, analyst, or data scientist who wants to gain experience in performing text data mining and analytics with R, then this book is for you. Exposure to working with statistical methods and language processing would be helpful. What You Will Learn Get acquainted with some of the highly efficient R packages such as OpenNLP and RWeka to perform various steps in the text mining process Access and manipulate data from different sources such as JSON and HTTP Process text using regular expressions Get to know the different approaches of tagging texts, such as POS tagging, to get started with text analysis Explore different dimensionality reduction techniques, such as Principal Component Analysis (PCA), and understand its implementation in R Discover the underlying themes or topics that are present in an unstructured collection of documents, using common topic models such as Latent Dirichlet Allocation (LDA) Build a baseline sentence completing application Perform entity extraction and named entity recognition using R In Detail Text Mining (or text data mining or text analytics) is the process of extracting useful and high-quality information from text by devising patterns and trends. R provides an extensive ecosystem to mine text through its many frameworks and packages. Starting with basic information about the statistics concepts used in text mining, this book will teach you how to access, cleanse, and process text using the R language and will equip you with the tools and the associated knowledge about different tagging, chunking, and entailment approaches and their usage in natural language processing. Moving on, this book will teach you different dimensionality reduction techniques and their implementation in R. Next, we will cover pattern recognition in text data utilizing classification mechanisms, perform entity recognition, and develop an ontology learning framework. By the end of the book, you will develop a practical application from the concepts learned, and will understand how text mining can be leveraged to analyze the massively available data on social media. Style and approach This book takes a hands-on, example-driven approach to the text mining process with lucid implementation in R.

Biplots in Practice

Over the last forty years there has been a growing interest to extend probability theory and statistics and to allow for more flexible modelling of imprecision, uncertainty, vagueness and ignorance. The fact that in many real-life situations data uncertainty is not only present in the form of randomness (stochastic uncertainty) but also in the form of imprecision/fuzziness is but one point underlining the need for a widening of statistical tools. Most such extensions originate in a "softening" of classical methods, allowing, in particular, to work with

imprecise or vague data, considering imprecise or generalized probabilities and fuzzy events, etc. About ten years ago the idea of establishing a recurrent forum for discussing new trends in the before-mentioned context was born and resulted in the first International Conference on Soft Methods in Probability and Statistics (SMPS) that was held in Warsaw in 2002. In the following years the conference took place in Oviedo (2004), in Bristol (2006) and in Toulouse (2008). In the current edition the conference returns to Oviedo. This edited volume is a collection of papers presented at the SMPS 2010 conference held in Mieres and Oviedo. It gives a comprehensive overview of current research into the fusion of soft methods with probability and statistics.

An Introduction to Applied Multivariate Analysis with R

Full of real-world case studies and practical advice, Exploratory Multivariate Analysis by Example Using R, Second Edition focuses on four fundamental methods of multivariate exploratory data analysis that are most suitable for applications. It covers principal component analysis (PCA) when variables are quantitative, correspondence analysis (CA) a

Exploratory Multivariate Analysis by Example Using R

ggplot2

Discovering knowledge from big multivariate data, recorded every days, requires specialized machine learning techniques. This book presents an easy to use practical guide in R to compute the most popular machine learning methods for exploring real word data sets, as well as, for building predictive models. The main parts of the book include: A) Unsupervised learning methods, to explore and discover knowledge from a large multivariate data set using clustering and principal component methods. You will learn hierarchical clustering, k-means, principal component analysis and correspondence analysis methods. B) Regression analysis, to predict a quantitative outcome value using linear regression and non-linear regression strategies. C) Classification techniques, to predict a qualitative outcome value using logistic regression, discriminant analysis, naive bayes classifier and support vector machines. D) Advanced machine learning methods, to build robust regression and classification models using k-nearest neighbors methods, decision tree models, ensemble methods (bagging, random forest and boosting). E) Model selection methods, to select automatically the best combination of predictor variables for building an optimal predictive model. These include, best subsets selection methods, stepwise regression and penalized regression (ridge, lasso and elastic net regression models). We also present principal component-based regression methods, which are useful when the data contain multiple correlated predictor variables. F) Model validation and evaluation techniques for measuring the performance of a predictive model. G) Model diagnostics for detecting and fixing a potential problems in a predictive model. The book presents the basic principles of these tasks and provide many examples in R. This book offers solid guidance in data mining for students and researchers. Key features: - Covers machine learning algorithm and implementation - Key

mathematical concepts are presented - Short, self-contained chapters with practical examples.

Quantitative Investment Portfolio Analytics in R

Over 100 hands-on recipes to effectively solve real-world data problems using the most popular R packages and techniques About This Book Gain insight into how data scientists collect, process, analyze, and visualize data using some of the most popular R packages Understand how to apply useful data analysis techniques in R for real-world applications An easy-to-follow guide to make the life of data scientist easier with the problems faced while performing data analysis Who This Book Is For This book is for those who are already familiar with the basic operation of R, but want to learn how to efficiently and effectively analyze real-world data problems using practical R packages. What You Will Learn Get to know the functional characteristics of R language Extract, transform, and load data from heterogeneous sources Understand how easily R can confront probability and statistics problems Get simple R instructions to quickly organize and manipulate large datasets Create professional data visualizations and interactive reports Predict user purchase behavior by adopting a classification approach Implement data mining techniques to discover items that are frequently purchased together Group similar text documents by using various clustering methods In Detail This cookbook offers a range of data analysis samples in simple and straightforward R code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently. The first section deals with how to create R functions to avoid the unnecessary duplication of code. You will learn how to prepare, process, and perform sophisticated ETL for heterogeneous data sources with R packages. An example of data manipulation is provided, illustrating how to use the "dplyr" and "data.table" packages to efficiently process larger data structures. We also focus on "ggplot2" and show you how to create advanced figures for data exploration. In addition, you will learn how to build an interactive report using the "ggvis" package. Later chapters offer insight into time series analysis on financial data, while there is detailed information on the hot topic of machine learning, including data classification, regression, clustering, association rule mining, and dimension reduction. By the end of this book, you will understand how to resolve issues and will be able to comfortably offer solutions to problems encountered while performing data analysis. Style and approach This easy-to-follow guide is full of hands-on examples of data analysis with R. Each topic is fully explained beginning with the core concept, followed by step-by-step practical examples, and concluding with detailed explanations of each concept used.

Statistics for Marketing and Consumer Research

R is a free, open source programming language that's become a popular standard for financial and economic analysis. Quantitative Investment Portfolio Analytics In R is your guide to getting started with modeling portfolio risk and return in R. Even if you have no experience with the software, you'll be fluent in R at a basic level after reading this short primer. The chapters provide step-by-step instructions for tapping into R's powerful capabilities for portfolio analytics.

Quantitative Methods in Archaeology Using R

A far-reaching course in practical advanced statistics for biologists using R/Bioconductor, data exploration, and simulation.

Modern Factor Analysis

Most data sets collected by researchers are multivariate, and in most cases, the variables need to be examined simultaneously to get the most informative results. This book covers the core multivariate methodology along with some basic theory for each method described. It also provides the necessary R and S-PLUS code for each analysis.

An R and S-Plus® Companion to Multivariate Analysis

Provides both rich theory and powerful applications Figures are accompanied by code required to produce them Full color figures

Statistics

Thoroughly revised and updated, this new edition of the text that helped define the field continues to present important methods in the quantitative analysis of geologic data, while showing students how statistics and computing can be applied to commonly encountered problems in the earth sciences. In addition to new and expanded coverage of key topics, the Third Edition features new pedagogy, end-of-chapter review exercises, and an accompanying website that contains all of the data for every example and exercise found in the book.

Statistics and Data Analysis in Geology

Discover what you can do with R! Introducing the R system, covering standard regression methods, then tackling more advanced topics, this book guides users through the practical, powerful tools that the R system provides. The emphasis is on hands-on analysis, graphical display, and interpretation of data. The many worked examples, from real-world research, are accompanied by commentary on what is done and why. The companion website has code and datasets, allowing readers to reproduce all analyses, along with solutions to selected exercises and updates. Assuming basic statistical knowledge and some experience with data analysis (but not R), the book is ideal for research scientists, final-year undergraduate or graduate-level students of applied statistics, and practising statisticians. It is both for learning and for reference. This third edition expands upon topics such as Bayesian inference for regression, errors in variables, generalized linear mixed models, and random forests.

Multiple Factor Analysis by Example Using R

Balancing simplicity with technical rigour, this practical guide to the statistical techniques essential to research in marketing and related fields, describes each method as well as showing how they are applied. The book is accompanied by two

real data sets to replicate examples and with exercises to solve, as well as detailed guidance on the use of appropriate software including: - 750 powerpoint slides with lecture notes and step-by-step guides to run analyses in SPSS (also includes screenshots) - 136 multiple choice questions for tests This is augmented by in-depth discussion of topics including: - Sampling - Data management and statistical packages - Hypothesis testing - Cluster analysis - Structural equation modelling

Modern Statistics for Modern Biology

A new edition of this popular text on robust statistics, thoroughly updated to include new and improved methods and focus on implementation of methodology using the increasingly popular open-source software R. Classical statistics fail to cope well with outliers associated with deviations from standard distributions. Robust statistical methods take into account these deviations when estimating the parameters of parametric models, thus increasing the reliability of fitted models and associated inference. This new, second edition of Robust Statistics: Theory and Methods (with R) presents a broad coverage of the theory of robust statistics that is integrated with computing methods and applications. Updated to include important new research results of the last decade and focus on the use of the popular software package R, it features in-depth coverage of the key methodology, including regression, multivariate analysis, and time series modeling. The book is illustrated throughout by a range of examples and applications that are supported by a companion website featuring data sets and R code that allow the reader to reproduce the examples given in the book. Unlike other books on the market, Robust Statistics: Theory and Methods (with R) offers the most comprehensive, definitive, and up-to-date treatment of the subject. It features chapters on estimating location and scale; measuring robustness; linear regression with fixed and with random predictors; multivariate analysis; generalized linear models; time series; numerical algorithms; and asymptotic theory of M-estimates. Explains both the use and theoretical justification of robust methods Guides readers in selecting and using the most appropriate robust methods for their problems Features computational algorithms for the core methods Robust statistics research results of the last decade included in this 2nd edition include: fast deterministic robust regression, finite-sample robustness, robust regularized regression, robust location and scatter estimation with missing data, robust estimation with independent outliers in variables, and robust mixed linear models. Robust Statistics aims to stimulate the use of robust methods as a powerful tool to increase the reliability and accuracy of statistical modelling and data analysis. It is an ideal resource for researchers, practitioners, and graduate students in statistics, engineering, computer science, and physical and social sciences.

Mathematical and Statistical Methods in Food Science and Technology

Although there are several good books on unsupervised machine learning, we felt that many of them are too theoretical. This book provides practical guide to cluster analysis, elegant visualization and interpretation. It contains 5 parts. Part I provides a quick introduction to R and presents required R packages, as well as, data formats and dissimilarity measures for cluster analysis and visualization. Part

II covers partitioning clustering methods, which subdivide the data sets into a set of k groups, where k is the number of groups pre-specified by the analyst. Partitioning clustering approaches include: K-means, K-Medoids (PAM) and CLARA algorithms. In Part III, we consider hierarchical clustering method, which is an alternative approach to partitioning clustering. The result of hierarchical clustering is a tree-based representation of the objects called dendrogram. In this part, we describe how to compute, visualize, interpret and compare dendrograms. Part IV describes clustering validation and evaluation strategies, which consists of measuring the goodness of clustering results. Among the chapters covered here, there are: Assessing clustering tendency, Determining the optimal number of clusters, Cluster validation statistics, Choosing the best clustering algorithms and Computing p-value for hierarchical clustering. Part V presents advanced clustering methods, including: Hierarchical k-means clustering, Fuzzy clustering, Model-based clustering and Density-based clustering.

Deep Learning with R

Mathematical and Statistical Approaches in Food Science and Technology offers an accessible guide to applying statistical and mathematical technologies in the food science field whilst also addressing the theoretical foundations. Using clear examples and case-studies by way of practical illustration, the book is more than just a theoretical guide for non-statisticians, and may therefore be used by scientists, students and food industry professionals at different levels and with varying degrees of statistical skill.

Robust Statistics

Principal Components Analysis

This book offers readers an accessible introduction to the world of multivariate statistics in the life sciences, providing a comprehensive description of the general data analysis paradigm, from exploratory analysis (principal component analysis, self-organizing maps and clustering) to modeling (classification, regression) and validation (including variable selection). It also includes a special section discussing several more specific topics in the area of chemometrics, such as outlier detection, and biomarker identification. The corresponding R code is provided for all the examples in the book; and scripts, functions and data are available in a separate R package. This second revised edition features not only updates on many of the topics covered, but also several sections of new material (e.g., on handling missing values in PCA, multivariate process monitoring and batch correction).

ROMANCE ACTION & ADVENTURE MYSTERY & THRILLER BIOGRAPHIES & HISTORY CHILDREN'S YOUNG ADULT FANTASY HISTORICAL FICTION HORROR LITERARY FICTION NON-FICTION SCIENCE FICTION