Analyzing Compositional Data With R Use R

This book presents the statistical analysis of compositional data using the log-ratio approach. It includes a wide range of classical and robust statistical methods adapted for compositional data analysis, such as supervised and unsupervised methods like PCA, correlation analysis, classification and regression. In addition, it considers special data structures like high-dimensional compositions and compositional tables. The methodology introduced is also frequently compared to methods which ignore the specific nature of compositional data. It focuses on practical aspects of compositional data analysis rather than on detailed theoretical derivations, thus issues like graphical visualization and preprocessing (treatment of missing values, zeros, outliers and similar artifacts) form an important part of the book. Since it is primarily intended for researchers and students from applied fields like geochemistry, chemometrics, biology and natural sciences, economics, and social sciences, all the proposed methods are accompanied by worked-out examples in R using the package robCompositions.

This new edition to the classic book by ggplot2 creator Hadley Wickham highlights compatibility with knitr and RStudio. ggplot2 is a data visualization package for R that helps users create data graphics, including those that are multi-layered, with ease. With ggplot2, it's easy to: produce handsome, publication-quality plots with automatic legends created from the plot specification superimpose multiple layers (points, lines, maps, tiles, box plots) from different data sources with automatically adjusted common scales add customizable smoothers that use powerful modeling capabilities of R, such as loess, linear models, generalized additive models, and robust regression save any ggplot2 plot (or part thereof) for later modification or reuse create custom themes that capture in-house or journal style requirements and that can easily be applied to multiple plots approach a graph from a visual perspective, thinking about how each component of the data is represented on the final plot This book will be useful to everyone who has struggled with displaying data in an informative and attractive way. Some basic knowledge of R is necessary (e.g., importing data into R). ggplot2 is a mini-language specifically tailored for producing graphics, and you'll learn everything you need in the book. After reading this book you'll be able to produce graphics customized precisely for your problems, and you'll find it easy to get graphics out of your head and on to the screen or page.

This guide for practicing statisticians, data scientists, and R users and programmers will teach the essentials of preprocessing: data leveraging the R programming language to easily and quickly turn noisy data into usable pieces of information. Data wrangling, which is also commonly referred to as data munging, transformation, manipulation, janitor work, etc., can be a painstakingly laborious process. Roughly 80% of data analysis is spent on cleaning and preparing data; however, being a prerequisite to the rest of the data analysis workflow (visualization, analysis, reporting), it is essential that one become fluent and efficient in data wrangling techniques. This book will guide the user through the data wrangling process via a step-by-step tutorial approach and provide a solid foundation for working with data in R. The author's goal is to teach the user how to easily wrangle data in order to spend more time on understanding the content of the data. By the end of the book, the user will have learned: How to work with different types of data such as numerics, characters, regular expressions, factors, and dates The difference between different data structures and how to create, add additional components to, and subset each data structure How to acquire and parse data from locations previously inaccessible How to develop functions and use loop control structures to reduce code redundancy How to use pipe operators to simplify code and make it more readable How to reshape the layout of data and manipulate, summarize, and join data sets

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

High-throughput sequencing technologies are widely used to study microbial ecology across species and habitats in order to understand the impacts of microbial communities on host health, metabolism, and the environment. Due to the dynamic nature of microbial communities, longitudinal microbiome analyses play an essential role in these types of investigations. Key questions in microbiome studies aim at identifying specific microbial taxa, enterotypes, genes, or metabolites associated with specific outcomes, as well as potential factors that influence microbial communities. However, the characteristics of microbiome data, such as sparsity and skewedness, combined with the nature of data collection, reflected often as uneven sampling or missing data, make commonly employed statistical approaches to handle repeated measures in longitudinal studies inadequate. Therefore, many researchers have begun to investigate methods that could improve incorporating these features when studying clinical, host, metabolic, or environmental associations with longitudinal microbiome data. In addition to the inferential aspect, it is also becoming apparent that visualization of high dimensional data in a way which is both intelligible and comprehensive is another difficult challenge that microbiome researchers face. Visualization is crucial in both the analysis and understanding of metagenomic data. Researchers must create clear graphic representations that give biological insight without being overly complicated. Thus, this Research Topic seeks to both review and provide novels approaches that are being developed to integrate microbiome data and complex metadata into meaningful mathematical, statistical and computational models. We believe this topic is fundamental to understanding the importance of microbial communities and provides a useful reference for other investigators approaching the field.

Modeling and Analysis of Compositional Data presents a practical and comprehensive introduction to the analysis of compositional data along with numerous examples to illustrate both theory and application of each method. Based upon short courses delivered by the authors, it provides a complete and current compendium of fundamental to advanced methodologies along with exercises at the end of each chapter to improve understanding, as well as data and a solutions manual which is available on an accompanying website. Complementing Pawlowsky-Glahn's earlier collective text that provides an overview of the state-of-the-art in this field, Modeling and Analysis of Compositional Data fills a gap in the literature for a much-needed manual for teaching, self learning or consulting.

This book is written for scientists and practitioners interested in deepening their knowledge of the sustainable production of bioenergy from wood in tropical and sub-tropical countries. Utilising the value chain concept, this book outlines the necessary aspects for managing sustainable bioenergy production. A wide range of topics is covered including biomass localization, modelling and upscaling, production management in woodlands and plantations, and transport and logistics. Biomass quality and conversion pathways are examined in order to match the conversion technology with the available biomass. A section is dedicated to issues surrounding sustainability. The issues, covered in a life-cycle assessment of the bioenergy system, include socio-economic challenges, local effects on water, biodiversity, nutrient-sustainability and global impacts. Through this holistic approach and supporting examples from tropical and sub-tropical countries, the reader

is guided in designing and implementing a value chain as the main management instrument for sustainable wood.

This book presents a collection of papers on topics in the field of strategic mine planning, including orebody modeling, mine-planning optimization and the optimization of mining complexes. Elaborating on the state of the art in the field, it describes the latest technologies and related research as well as the applications of a range of related technologies in diverse industrial contexts.

Analyzing Compositional Data with RSpringer Science & Business Media

Practical, example-driven introduction to maximum likelihood for the social sciences. Emphasizes computation in R, model selection and interpretation.

With research from leading scientists, Sedentary Behavior and Health: Concepts, Assessments, and Intervention presents evidence on sedentary behavior, its apparent health risks, and suggestions on measuring and altering this behavior.

This textbook describes the broadening methodology spectrum of psychological measurement in order to meet the statistical needs of a modern psychologist. The way statistics is used, and maybe even perceived, in psychology has drastically changed over the last few years; computationally as well as methodologically. R has taken the field of psychology by storm, to the point that it can now safely be considered the lingua franca for statistical data analysis in psychology. The goal of this book is to give the reader a starting point when analyzing data using a particular method, including advanced versions, and to hopefully motivate him or her to delve deeper into additional literature on the method. Beginning with one of the oldest psychometric model formulations, the true score model, Mair devotes the early chapters to exploring confirmatory factor analysis, modern test theory, and a sequence of multivariate exploratory method. Subsequent chapters present special techniques useful for modern psychological applications including correlation networks, sophisticated parametric clustering techniques, longitudinal measurements on a single participant, and functional magnetic resonance imaging (fMRI) data. In addition to using real-life data sets to demonstrate each method, the book also reports each method in three parts-- first describing when and why to apply it, then how to compute the method in R, and finally how to present, visualize, and interpret the results. Requiring a basic knowledge of statistical methods and R software, but written in a casual tone, this text is ideal for graduate students in psychology. Relevant courses include methods of scaling, latent variable modeling, psychometrics for graduate students in Psychology, and multivariate methods in the social sciences.

This book is the first single source volume to fully address this prevalent practice in both its analytical and modeling aspects. The information discussed presents the use of data consisting of rankings in such diverse fields as psychology, animal science, educational testing, sociology, economics, and biology. This book systematically presents the basic models and methods for analyzing data in the form of ranks. Integrating material from a wide range of fields, this book applies graphical, numerical, and modeling techniques to data sets, uncovering fascinating structures in the rank data. Providing the most extensive coverage of the subject found in statistical literature, this book will be a welcomed reference to statisticians. In addition, this volume is also accessible to people in all areas of quantitative research. Researchers in psychology and consumer preference will discover a valuable resource; and sociologists, biologists, political and animal scientists will also benefit. As a text, it will be ideal for graduate students in courses on statistics and other quantitative disciplines.

Petra in modern southern Jordan is universally known as one of the most frequented touristic sites in the Near East, inscribed on the UNESCO world heritage list. Modern visitors are attracted by the romantic aspect of the rock-cut tomb façades, heavily contrasting in their baroque stile with the desert like surrounding of the rocky and arid landscape. These monuments were the result of the long time presence of the Arab tribe of the Nabataeans who made Petra their capital when they became, at least partially, sedentarised during the Hellenistic period, i.e between the late 4th to late 1st centuries BCE. How exactly this process of sedentarisation happened, how the site of Petra changed from a temporary dwelling place of a small Bedouin tribe to one of the blinking capitals of the ancient Near East that attracted - as it is the case today - visitors from all over the world, was the subject of a three years research program, jointly sponsored by the French Agence Nationale de la Recherche (ANR) and the German Research Foundation (DFG). At the end of the program, an international conference, held in Berlin in December 2011, brought together several dozen of scholars from all over the world in order to pinpoint the state of research on the Formation of the Nabataean capital. The contributions of the present volume focus on questions related to the natural environment of the site, on the geology and geography as well as on architecture, small finds and social dynamics, probably the clue for a better understanding of the functioning of the Nabataean kingdom and its capital Petra.

This book presents the statistical analysis of compositional data sets, i.e., data in percentages, proportions, concentrations, etc. The subject is covered from its grounding principles to the practical use in descriptive exploratory analysis, robust linear models and advanced multivariate statistical methods, including zeros and missing values, and paying special attention to data visualization and model display issues. Many illustrated examples and code chunks guide the reader into their modeling and interpretation. And, though the book primarily serves as a reference guide for the R package "compositions," it is also a general introductory text on Compositional Data Analysis. Awareness of their special characteristics spread in the Geosciences in the early sixties, but a strategy for properly dealing with them was not available until the works of Aitchison in the eighties. Since then, research has expanded our understanding of their theoretical principles and the potentials and limitations of their interpretation. This is the first comprehensive textbook addressing these issues, as well as their practical implications with regard to software. The book is intended for scientists interested in statistically analyzing their compositional data. The subject enjoys relatively broad awareness in the geosciences and environmental sciences, but the spectrum of recent applications also covers areas like medicine, official statistics, and economics. Readers should be familiar with basic univariate and multivariate statistics. Knowledge of R is recommended but not required, as the book is self-contained.

This unique book addresses the statistical modelling and analysis of microbiome data using cutting-edge R software. It includes real-world data from the authors' research and from the public domain, and discusses the implementation of R for data analysis step by step. The data and R computer programs are publicly available, allowing readers to replicate the model development and data analysis presented in each chapter, so that these new methods can be readily applied in their own research. The book also discusses recent developments in statistical modelling and data analysis in microbiome research, as well as the latest advances in next-generation sequencing and big data in methodological development and applications. This timely book will greatly benefit all readers involved in microbiome, ecology and microarray data analyses, as well as other fields of research.

This dictionary includes a number of mathematical, statistical and computing terms and their definitions to assist geoscientists and provide guidance on the methods and terminology encountered in the literature. Each technical term used in the explanations can be found in the dictionary which also includes explanations of basics, such as trigonometric functions and logarithms. There are also citations from the relevant literature to show the term's first use in mathematics, statistics, etc. and its subsequent usage in geosciences.

The aim of this book is to unlock the power of the freeware R language to advanced university students and researchers dealing with whole-rock geochemistry of (meta-) igneous rocks. The first part covers data input/output, calculation of commonly used indexes and plotting in R. The core of the book then focusses on the presentation and practical implementations of modelling techniques used for fingerprinting processes such as partial melting, fractional crystallization, binary mixing or AFC using major-, trace-element and radiogenic isotope data. The reader will be given a

firm theoretical basis for forward/reverse modelling, followed by exercises dealing with typical problems likely to be encountered in real life, and their solutions using R. The concluding sections demonstrate, using practical examples, how a researcher can proceed in developing a realistic model simulating natural systems. The appendices outline the fundamentals of the R language and provide a quick introduction to the open-source R-package GCDkit for interpretation of whole-rock geochemical data from igneous and metamorphic rocks.

The book provides a description of the process of health economic evaluation and modelling for cost-effectiveness analysis, particularly from the perspective of a Bayesian statistical approach. Some relevant theory and introductory concepts are presented using practical examples and two running case studies. The book also describes in detail how to perform health economic evaluations using the R package BCEA (Bayesian Cost-Effectiveness Analysis). BCEA can be used to post-process the results of a Bayesian cost-effectiveness model and perform advanced analyses producing standardised and highly customisable outputs. It presents all the features of the package, including its many functions and their practical application, as well as its user-friendly web interface. The book is a valuable resource for statisticians and practitioners working in the field of health economics wanting to simplify and standardise their workflow, for example in the preparation of dossiers in support of marketing authorisation, or academic and scientific publications.

How best to interpret and apply geochemical data to understand geological processes, for graduate students, researchers, and professionals.

Sediment provenance studies concern the origin, composition, transportation and deposition of detritus and therefore are an important part of understanding the links between basinal sedimentation, and hinterland tectonics and unroofing. Such studies can add value at many stages of hydrocarbon exploitation, from identifying regional-scale crustal affinities and sediment dispersal patterns during the earliest stages of exploration, to detailed correlation in producing reservoirs and understanding the impact of mineralogy on reservoir diagenesis. The volume showcases the wide variety of techniques available, using examples and applications from all aspects of sediment provenance research. The papers are organized into four sets around the following themes: • Overview: applications of provenance information in hydrocarbon reservoir sandstones • Provenance, diagenesis and reservoir quality • Provenance studies linking sediment to source • Looking forward: development of techniques and data handling This book is dedicated to the memory of Maria Mange and Robert A. Scott.

This book is a complete introduction to the power of R for marketing research practitioners. The text describes statistical models from a conceptual point of view with a minimal amount of mathematics, presuming only an introductory knowledge of statistics. Hands-on chapters accelerate the learning curve by asking readers to interact with R from the beginning. Core topics include the R language, basic statistics, linear modeling, and data visualization, which is presented throughout as an integral part of analysis. Later chapters cover more advanced topics yet are intended to be approachable for all analysts. These sections examine logistic regression, customer segmentation, hierarchical linear modeling, market basket analysis, structural equation modeling, and conjoint analysis in R. The text uniquely presents Bayesian models with a minimally complex approach, demonstrating and explaining Bayesian methods alongside traditional analyses for analysis of variance, linear models, and metric and choice-based conjoint analysis. With its emphasis on data visualization, model assessment, and development of statistical intuition, this book provides guidance for any analyst looking to develop or improve skills in R for marketing applications.

This innovative text is the first to map regional grammatical variation in written Standard American English.

This book introduces readers to the basic concepts of Heart Rate Variability (HRV) and its most important analysis algorithms using a hands-on approach based on the open-source RHRV software. HRV refers to the variation over time of the intervals between consecutive heartbeats. Despite its apparent simplicity, HRV is one of the most important markers of the autonomic nervous system activity and it has been recognized as a useful predictor of several pathologies. The book discusses all the basic HRV topics, including the physiological contributions to HRV, clinical applications, HRV data acquisition, HRV data manipulation and HRV analysis using time-domain, frequency-domain, time-frequency, nonlinear and fractal techniques. Detailed examples based on real data sets are provided throughout the book to illustrate the algorithms and discuss the physiological implications of the results. Offering a comprehensive guide to analyzing beat information with RHRV, the book is intended for masters and Ph.D. students in various disciplines such as biomedical engineering, human and veterinary medicine, biology, and pharmacy, as well as researchers conducting heart rate variability analyses on both human and animal data.

A comprehensive treatment of statistical applications for solvingreal-world environmental problems A host of complex problems face today's earth science community,such as evaluating the supply of remaining non-renewable energyresources, assessing the impact of people on the environment,understanding climate change, and managing the use of water. Propercollection and analysis of data using statistical techniquescontributes significantly toward the solution of these problems.Statistics for Earth and Environmental Scientists presentsimportant statistical concepts through data analytic tools andshows readers how to apply them to real-world problems. The authors present several different statistical approaches tothe environmental sciences, including Bayesian and nonparametricmethodologies. The book begins with an introduction to types ofdata, evaluation of data, modeling and estimation, randomvariation, and sampling—all of which are explored throughcase studies that use real data from earth science applications.Subsequent chapters focus on principles of modeling and the keymethods and techniques for analyzing scientific data,including: Interval estimation and Methods for analyzinghypothesis testingof means time series data Spatial statistics Multivariate analysis Discrete distributions Experimental design Most statistical models are introduced by concept andapplication, given as equations, and then accompanied by heuristicjustification rather than a formal proof. Data analysis, modelbuilding, and statistical inference are stressed throughout, andreaders are encouraged to collect their own data to incorporateinto the exercises at the end of each chapter. Most data sets,graphs, and analyses are computed using R, but can be worked withusing any statistical computing software. A related websitefeatures additional data sets, answers to selected exercises, and Rcode for the

book's examples. Statistics for Earth and Environmental Scientists is anexcellent book for courses on quantitative methods in geology,geography, natural resources, and environmental sciences at theupper-undergraduate and graduate levels. It is also a valuablereference for earth scientists, geologists, hydrologists, andenvironmental statisticians who collect and analyze data in theireveryday work.

This book constitutes the thoroughly refereed proceedings of the Fourth International Conference on Mathematics and Computation in Music, MCM 2013, held in Montreal, Canada, in June 2013. The 18 papers presented were carefully reviewed and selected from numerous submissions. They are promoting the collaboration and exchange of ideas among researchers in music theory, mathematics, computer science, musicology, cognition and other related fields.

The principal aim of this book is to provide a wide range of information and a useful reference for researchers interested to investigate heavy mineral assemblages in different geological settings and for a variety of purposes. The methodological developments achieved in recent years for the identification of heavy minerals in a wide grain-size range are illustrated. All factors that affect heavy mineral concentration and relative proportions, including hydraulic sorting, mechanical abrasion, chemical weathering, and post-depositional dissolution, and all factors able to introduce analytical, environmental, or diagenetic bias are thoroughly addressed. A proper integration of multiple techniques including bulk sediment, multi-mineral, and single-mineral methods are discussed by renowned authors in their invited contributions.

The first step-by-step guide to the quantitative analysis of archaeological data using the R statistical computing system. Environmental engineering has a leading role in the elimination of ecological threats, and deals, in brief, with securing technically the conditions which create a safe environment for mankind to live in. Due to its interdisciplinary character it can deal with a wide range of technical and technological problems. Since environmental engineering use

This book and app is for practitioners, professionals, researchers, and students who want to learn how to make a plot within the R environment using ggplot2, step-by-step without coding. In widespread use in the statistical communities, R is a free software language and environment for statistical programming and graphics. Many users find R to have a steep learning curve but to be extremely useful once overcome. ggplot2 is an extremely popular package tailored for producing graphics within R but which requires coding and has a steep learning curve itself, and Shiny is an open source R package that provides a web framework for building web applications using R without requiring HTML, CSS, or JavaScript. This manual—"integrating" R, ggplot2, and Shiny—introduces a new Shiny app, Learn ggplot2, that allows users to make plots easily without coding. With the Learn ggplot2 Shiny app, users can make plots using ggplot2 without having to code each step, reducing typos and error messages and allowing users to become familiar with ggplot2 code. The app makes it easy to apply themes, make multiplots (combining several plots into one plot), and download plots as PNG, PDF, or PowerPoint files with editable vector graphics. Users can also make plots on any computer or smart phone. Learn ggplot2 Using Shiny App allows users to Make publication-ready plots in minutes without coding Download plots with desired width, height, and resolution Plot and download plots in png, pdf, and PowerPoint formats, with or without R code and with editable vector graphics

????:???????????????????????????????????????

1. Introduction. 2. Regionalized Compositions. 3. Spatial Covariance Structure. 4. Concepts of Null Correlation. 5. Cokriging. 6. Practical Aspects of Compositional Data Analysis. 7. Application to Real Data. Summary and Prospects. References. Index.

The YUIMA package is the first comprehensive R framework based on S4 classes and methods which allows for the simulation of stochastic differential equations driven by Wiener process, Lévy processes or fractional Brownian motion, as well as CARMA, COGARCH, and Point processes. The package performs various central statistical analyses such as quasi maximum likelihood estimation, adaptive Bayes estimation, structural change point analysis, hypotheses testing, asynchronous covariance estimation, lead-lag estimation, LASSO model selection, and so on. YUIMA also supports stochastic numerical analysis by fast computation of the expected value of functionals of stochastic processes through automatic asymptotic expansion by means of the Malliavin calculus. All models can be multidimensional, multiparametric or non parametric.The book explains briefly the underlying theory for simulation and inference of several classes of stochastic processes and then presents both simulation experiments and applications to real data. Although these processes have been originally proposed in physics and more recently in finance, they are becoming popular also in biology due to the fact the time course experimental data are now available. The YUIMA package, available on CRAN, can be freely downloaded and this companion book will make the user able to start his or her analysis from the first page.

It is difficult to imagine that the statistical analysis of compositional data has been a major issue of concern for more than 100 years. It is even more difficult to realize that so many statisticians and users of statistics are unaware of the particular problems affecting compositional data, as well as their solutions. The issue of ``spurious correlation'', as the situation was phrased by Karl Pearson back in 1897, affects all data that measures parts of some whole, such as percentages, proportions, ppm and ppb. Such measurements are present in all fields of science, ranging from geology, biology, environmental sciences, forensic sciences, medicine and hydrology. This book presents the history and development of compositional data analysis along with Aitchison's log-ratio approach. Compositional Data Analysis describes the state of the art both in theoretical fields as well as applications in the different fields of science. Key Features: Reflects the state-of-the-art in compositional data analysis. Gives an overview of the historical development of compositional data analysis, as well as basic concepts and procedures. Looks at advances in algebra and calculus on the simplex. Presents applications in different fields of science, including, genomics, ecology, biology, geochemistry, planetology, chemistry and economics. Explores connections to correspondence analysis and the Dirichlet distribution. Presents a summary of three available software packages for compositional data analysis. Supported by an accompanying website featuring R code. Applied scientists working on compositional data analysis in any field of science, both in academia and professionals will benefit from this book, along with graduate students in any field of science working with compositional data.

Taking a multidisciplinary approach, this comprehensive Handbook comprises contributions from international researchers of diverse educational and research backgrounds. Chapters present methodological issues within marketing research, sharing the researchers' experiences of what does and does not work, as well as discussing challenges and avenues for innovation.

In this volume recent advances in the use of modern quantitative models for the analysis of various problems related to the dynamics of social and economic systems are presented. The majority chapters describe tools and techniques of broadly perceived computational intelligence, notably fuzzy logic, evolutionary computation, neural networks and some non-standard probabilistic and statistical analyses. Due to the high complexity of the systems and problems considered, in many situations it is necessary to consider at the same time analytic, topological and statistical aspects and apply appropriate procedures and algorithms. This volume is a direct result of vivid discussions held during the Fifth International Workshop on Dynamics of Social and Economical Systems (DYSES) which was held at Benevento, Italy September 20-25, 2010, as well as a couple of post-workshop meetings and consultations.

Copyright: 9f00ca8034a250f8e42ab1dda2c4ac0f