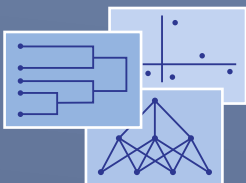


Studies in Classification, Data Analysis,
and Knowledge Organization

Leonardo Grilli · Monia Lupporelli ·
Carla Rampichini · Emilia Rocco ·
Maurizio Vichi *Editors*

Statistical Models and Methods for Data Science



 Springer

Studies in Classification, Data Analysis, and Knowledge Organization

Managing Editors

Wolfgang Gaul, Karlsruhe, Germany

Maurizio Vichi, Rome, Italy

Claus Weihs, Dortmund, Germany

Editorial Board

Daniel Baier, Bayreuth, Germany

Frank Critchley, Milton Keynes, UK

Reinhold Decker, Bielefeld, Germany

Edwin Diday, Paris, France

Michael Greenacre, Barcelona, Spain

Carlo Natale Lauro, Naples, Italy

Jacqueline Meulman, Leiden, The
Netherlands

Paola Monari, Bologna, Italy

Shizuhiko Nishisato, Toronto, Canada

Noboru Ohsumi, Tokyo, Japan

Otto Opitz, Augsburg, Germany

Gunter Ritter, Passau, Germany

Martin Schader, Mannheim, Germany

Studies in Classification, Data Analysis, and Knowledge Organization is a book series which offers constant and up-to-date information on the most recent developments and methods in the fields of statistical data analysis, exploratory statistics, classification and clustering, handling of information and ordering of knowledge. It covers a broad scope of theoretical, methodological as well as application-oriented articles, surveys and discussions from an international authorship and includes fields like computational statistics, pattern recognition, biological taxonomy, DNA and genome analysis, marketing, finance and other areas in economics, databases and the internet. A major purpose is to show the intimate interplay between various, seemingly unrelated domains and to foster the cooperation between mathematicians, statisticians, computer scientists and practitioners by offering well-based and innovative solutions to urgent problems of practice.

Leonardo Grilli · Monia Lupparelli ·
Carla Rampichini · Emilia Rocco · Maurizio Vichi
Editors

Statistical Models and Methods for Data Science

 Springer

Editors

Leonardo Grilli
Department of Statistics, Computer
Science, Applications “G. Parenti”
University of Florence
Florence, Italy

Monia Lupparelli
Department of Statistics, Computer
Science, Applications “G. Parenti”
University of Florence
Florence, Italy

Carla Rampichini
Department of Statistics, Computer
Science, Applications “G. Parenti”
University of Florence
Florence, Italy

Emilia Rocco
Department of Statistics, Computer
Science, Applications “G. Parenti”
University of Florence
Florence, Italy

Maurizio Vichi
Department of Statistical Sciences
Sapienza University of Rome
Rome, Italy

ISSN 1431-8814

ISSN 2198-3321 (electronic)

Studies in Classification, Data Analysis, and Knowledge Organization

ISBN 978-3-031-30163-6

ISBN 978-3-031-30164-3 (eBook)

<https://doi.org/10.1007/978-3-031-30164-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book offers a collection of papers focusing on methods and models in classification and data analysis. Several research topics are covered, ranging from statistical inference and modeling to clustering and factorial methods, from directional data analysis to time series analysis and small area estimation. Applications deal with new investigations in a variety of relevant fields: medicine, finance, engineering, marketing, and cyber risk, to cite a few.

These contributions are a selection of post-conference papers presented at the 13th meeting of the CLAssification and Data Analysis Group (CLADAG) of the Italian Statistical Society (SIS), organized by the Department of Statistics, Computer Science, Applications “G. Parenti” of the University of Florence (Italy) on September 9–11, 2021. The submitted papers followed a careful review process involving two reviewers per paper. In the end, 14 papers were selected for publication in this volume.

Due to the persistent uncertainty about the COVID-19 epidemic, CLADAG 2021 was entirely online. Despite this unfortunate situation, the Conference was highly participated and vital, as the wide range of contributions here collected shows.

CLADAG, a member of the International Federation of Classification Societies (IFCS), organizes an international scientific meeting every two years devoted to presenting theoretical and applied papers in classification and, more generally, data analysis. The meeting includes advanced methodological research in multivariate statistics, mathematical and statistical investigations, survey papers on the state of the art, real case studies, papers on numerical and algorithmic aspects, and applications in special fields of interest at the interface between classification and data science. The Conference aims to encourage the interchange of ideas in the mentioned research fields and disseminate new findings. CLADAG conferences, initiated in 1997 in Pescara (Italy), were soon considered an attractive information exchange market and became an important meeting point for people interested in classification and data analysis. Traditionally, a selection of the presented papers, fully peer-reviewed, is published in a volume of post-conference proceedings.

The Scientific Committee of the 2021 edition has planned Plenary and Invited Sessions to provide a fresh perspective on the state of the art of knowledge and research in the field. The scientific program of CLADAG 2021 is especially rich.

All in all, it comprises five Keynote Lectures, 26 Invited Sessions promoted by the members of the Scientific Program Committee, 10 Contributed Sessions, and a Plenary Session on Statistical Issues in the COVID-19 pandemic.

We thank all the session organizers for inviting renowned speakers from many countries. We are greatly indebted to the referees for the time spent carefully reviewing the papers collected in this book. Special thanks are due to the members of the Local Organizing Committee and all the people who collaborated with CLADAG 2021. Last but not least, we thank all the authors and participants, without whom the Conference would not have been possible. Above all, we thank all participants who chose this book to share their research findings. We hope this book will contribute to fostering new knowledge in the field.

Florence, Italy
November 2022

Leonardo Grilli
Monia Lupparelli
Carla Rampichini
Emilia Rocco
Maurizio Vichi

Contents

Clustering Financial Time Series by Dependency	1
Andrés M. Alonso, Carolina Gamboa, and Daniel Peña	
The Homogeneity Index as a Measure of Interrater Agreement for Ratings on a Nominal Scale	15
Giuseppe Bove	
Hierarchical Clustering of Income Data Based on Share Densities	27
Francesca Condino	
Optimal Coding of High-Cardinality Categorical Data in Machine Learning	39
Agostino Di Ciaccio	
Bayesian Multivariate Analysis of Mixed Data	53
Chiara Galimberti, Federico Castelletti, and Stefano Peluso	
Marginals Matrix Under a Generalized Mallows Model Based on the Power Divergence	67
Maria Kateri and Nikolay I. Nikolov	
Time Series Clustering Based on Forecast Distributions: An Empirical Analysis on Production Indices for Construction	81
Michele La Rocca, Francesco Giordano, and Cira Perna	
Partial Reconstruction of Measures from Halfspace Depth	93
Petra Laketa and Stanislav Nagy	
Posterior Predictive Assessment of IRT Models via the Hellinger Distance: A Simulation Study	107
Mariagiulia Matteucci and Stefania Mignani	
Shapley-Lorenz Values for Credit Risk Management	121
Niklas Bussmann, Roman Enzmann, Paolo Giudici, and Emanuela Raffinetti	

A Study of Lack-of-Fit Diagnostics for Models Fit to Cross-Classified Binary Variables 133
Maduranga Dassanayake and Mark Reiser

Robust Response Transformations for Generalized Additive Models via Additivity and Variance Stabilization 147
Marco Riani, Anthony C. Atkinson, and Aldo Corbellini

A Random-Coefficients Analysis with a Multivariate Random-Coefficients Linear Model 161
Laura Marcis, Maria Chiara Pagliarella, and Renato Salvatore

Parsimonious Mixtures of Matrix-Variate Shifted Exponential Normal Distributions 177
Salvatore D. Tomarchio, Luca Bagnato, and Antonio Punzo

Author Index 187