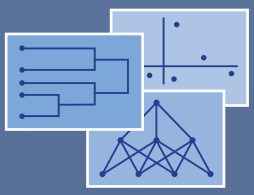
Studies in Classification, Data Analysis, and Knowledge Organization

Simona Balzano · Giovanni C. Porzio · Renato Salvatore · Domenico Vistocco · Maurizio Vichi *Editors*

Statistical Learning and Modeling in Data Analysis

Methods and Applications



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.

More information about this series at http://www.springer.com/series/1564

Simona Balzano · Giovanni C. Porzio ·
Renato Salvatore · Domenico Vistocco ·
Maurizio Vichi
Editors

Statistical Learning and Modeling in Data Analysis

Methods and Applications

Springer

Editors
Simona Balzano
Department of Economics and Law
University of Cassino and Southern Lazio
Cassino, Italy

Renato Salvatore
Department of Economics and Law
University of Cassino and Southern Lazio
Cassino, Italy

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

Giovanni C. Porzio
Department of Economics and Law
University of Cassino and Southern Lazio
Cassino, Italy

Domenico Vistocco
Department of Political Science
University of Naples Federico II
Naples, Italy

ISSN 1431-8814 ISSN 2198-3321 (electronic)
Studies in Classification, Data Analysis, and Knowledge Organization
ISBN 978-3-030-69943-7 ISBN 978-3-030-69944-4 (eBook)
https://doi.org/10.1007/978-3-030-69944-4

Mathematics Subject Classification: 62-06, 62-07, 62Fxx, 62Gxx, 62Hxx, 62Jxx, 62Kxx

© Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are reserved 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 for statistical learning and modeling in data analysis. A series of interesting applications are offered as well. 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 analyses within a variety of fields of interest: medicine, finance, engineering, marketing, cyber risk, to cite a few.

The book arises as post-proceedings of the 12th meeting of the CLAssification and Data Analysis Group (CLADAG) of the Italian Statistical Society (SIS), held in Cassino (IT), on September 11–13, 2019. The first CLADAG meeting was held in 1997, in Pescara (IT). CLADAG is also a member of the International Federation of Classification Societies (IFCS), founded in 1985. CLADAG promotes advanced methodological research in multivariate statistics with a special vocation towards Data Analysis and Classification. It supports the interchange of ideas in these fields of research, including the dissemination of concepts, numerical methods, algorithms, computational and applied results. This book is thus in line with the main CLADAG goals.

Thanks to the participation of renowned speakers, coming from 28 different countries, the scientific program of the CLADAG 2019 Conference was particularly engaging. It saw 5 Keynote Lectures, 32 Invited Sessions, 16 Contributed Sessions, a Round Table, and a Data Competition. The richness of the Conference program, and hence of this book, is definitely due to the Conference Scientific Committee and particularly to its Chair Francesca Greselin. We are indebted to their work. We are also indebted to the anonymous referees. They did a great job and helped us to improve the overall quality of this book.

Our gratitude also goes to the staff of the Department of Economics and Law, University of Cassino and Southern Lazio, who supported the conference and contributed to its success. A special thank goes to Livia Iannucci, who worked side by side with the Local Organizing Committee offering her precious administrative support before, during, and after the conference.

Above all, we are thankful to all the participants and to those who, among them, have chosen this book to share their research findings. Our wish is that this book will contribute to foster the creation of new knowledge in the field.

Cassino, Italy
26 November 2020

Simona Balzano
Giovanni C. Porzio
Renato Salvatore
Domenico Vistocco
Maurizio Vichi

Contents

Interpreting Effects in Generalized Linear Modeling 1
Alan Agresti, Claudia Tarantola, and Roberta Varriale

ACE, AVAS and Robust Data Transformations 9
Anthony C. Atkinson, Marco Riani, Aldo Corbellini, and Gianluca Morelli

On Predicting Principal Components Through Linear Mixed Models 17
Simona Balzano, Maja Bozic, Laura Marcis, and Renato Salvatore

Robust Model-Based Learning to Discover New Wheat Varieties and Discriminate Adulterated Kernels in X-Ray Images 29
Andrea Cappozzo, Francesca Greselin, and Thomas Brendan Murphy

A Dynamic Model for Ordinal Time Series: An Application to Consumers' Perceptions of Inflation 37
Marcella Corduas

Deep Learning to Jointly Analyze Images and Clinical Data for Disease Detection 47
Federica Crobu and Agostino Di Ciaccio

Studying Affiliation Networks Through Cluster CA and Blockmodeling 57
Daniela D'Ambrosio, Marco Serino, and Giancarlo Ragozini

Sectioning Procedure on Geostatistical Indices Series of Pavement Road Profiles 69
Mauro D'Apuzzo, Rose-Line Spacagna, Azzurra Evangelisti, Daniela Santilli, and Vittorio Nicolosi

Directional Supervised Learning Through Depth Functions: An Application to ECG Waves Analysis 79
Houyem Demni

Penalized Versus Constrained Approaches for Clusterwise Linear Regression Modeling 89
Roberto Di Mari, Stefano Antonio Gattone, and Roberto Rocci

Effect Measures for Group Comparisons in a Two-Component Mixture Model: A Cyber Risk Analysis 97
Maria Iannario and Claudia Tarantola

A Cramér–von Mises Test of Uniformity on the Hypersphere 107
Eduardo García-Portugués, Paula Navarro-Esteban, and Juan Antonio Cuesta-Albertos

On Mean And/or Variance Mixtures of Normal Distributions 117
Sharon X. Lee and Geoffrey J. McLachlan

Robust Depth-Based Inference in Elliptical Models 129
Stanislav Nagy and Jiří Dvořák

Latent Class Analysis for the Derivation of Marketing Decisions: An Empirical Study for BEV Battery Manufacturers 139
Friederike Paetz

Small Area Estimation Diagnostics: The Case of the Fay–Herriot Model 149
Maria Chiara Pagliarella

A Comparison Between Methods to Cluster Mixed-Type Data: Gaussian Mixtures Versus Gower Distance 163
Monia Ranalli and Roberto Rocci

Exploring the Gender Gap in Erasmus Student Mobility Flows 173
Marialuisa Restaino, Ilaria Primerano, and Maria Prosperina Vitale