Springer Proceedings in Mathematics & Statistics

Regina Bispo
Lígia Henriques-Rodrigues
Russell Alpizar-Jara
Miguel de Carvalho *Editors*

Recent Developments in Statistics and Data Science

SPE2021, Évora, Portugal, October 13–16

Springer

Springer Proceedings in Mathematics & Statistics

Volume 398

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Regina Bispo · Lígia Henriques-Rodrigues ·
Russell Alpizar-Jara · Miguel de Carvalho
Editors

Recent Developments in Statistics and Data Science

SPE2021, Évora, Portugal, October 13–16

Springer

Editors
Regina Bispo
NOVA School of Science
and Technology
Caparica, Portugal

Russell Alpizar-Jara
University of Évora
Évora, Portugal

Lígia Henriques-Rodrigues
University of Évora
Évora, Portugal

Miguel de Carvalho
University of Edinburgh
Edinburgh, UK

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-031-12765-6 ISBN 978-3-031-12766-3 (eBook)
https://doi.org/10.1007/978-3-031-12766-3

Mathematics Subject Classification: 62-06, 62P10, 62P25, 62R07

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
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

Organization

SPE 2021 was organized by the University of Évora and by the Portuguese Statistical Society (SPE).

Executive Commitee

Russell Alpizar-Jara (President), University of Évora (PT)
Dulce Gomes, University of Évora (PT)
Lígia Henriques-Rodrigues, University of Évora (PT)
Patrícia A. Filipe, ISCTE, University Institute of Lisbon (PT)

Scientific Committee

Miguel de Carvalho (President), University of Edinburgh (UK)
Fátima Ferreira, University of Trás-os-monte e Alto Douro (PT)
João Andrade e Silva, University of Lisbon (PT)
Luís Meira-Machado, University of Minho (PT)
Marco Costa, University of Aveiro (PT)
Maria Eduarda Silva, University of Oporto (PT)
Marília Antunes, University of Lisbon (PT)
Paula Brito, University of Oporto (PT)
Regina Bispo, Nova University of Lisbon (PT)
Rosário Oliveira, University of Lisbon (PT)
Russell Alpizar-Jara, University of Évora (PT)

Partners and Sponsoring Institutions

International Statistical Associations:

- Bernoulli Society
- Brazilian Statistical Association
- CWS (Caucus for Women in Statistics)
- FENStatS (Federation of the European National Statistical Societies)
- ISBA (International Society for Bayesian Analysis)
- RBras (Brazilian Region International Biometric Society)
- SGaPEIO (Sociedade Galega para a promoción da Estatística e da Investigación de Operacions)

National Statistical Associations:

- CLAD (Portuguese Association for Classification and Data Analysis)

Industry and Official Statistics:

- GADES (Data Analysis Solutions)
- INE (Statistics Portugal)
- PSE—Your Data Specialists
- PORDATA—Estatísticas sobre Portugal e Europa

Referees for Proceedings

Ana Freitas
Andy Lynch
Carlos A. Braumann
Clara Cordeiro
Dulce Gomes
Filipe Marques
Inês Sousa
Isabel Pereira
Jessica Silva Lomba
Kamil Turkman
Lisete Sousa
Luís Machado
M. Filomena Teodoro
Manuela Neves
Marco Costa
Maria Antónia Turkman
Maria de Fátima Ferreira
Maria Ivette Gomes
Maria Polidoro

Marília Antunes
Maurizio Sanarico
Nuno Sepúlveda
Patrícia de Zea Bermudez
Patrícia Filipe
Paula Brito
Paulo C. Rodrigues
Paulo M. M. Rodrigues
Pedro Campos
Raquel Menezes
Rita Sousa
Soraia Pereira
Tiago Marques
Vanda Lourenço

Welcome Message from the Editors

Dear authors, referees, and readers of

Recent Developments in Statistics and Data Science,

It is a great pleasure to welcome you to the proceedings of the XXV Congress of the Portuguese Statistical Society—**SPE 2021**—held during 13–16 October 2021 at Évora, Portugal. This was the first-time, ever, online SPE conference, and it gathered more than 200 delegates from all over the world.

The meeting was hosted by University of Évora, Portugal, in collaboration with the Portuguese Statistical Society, and we had a fantastic program including 4 plenary lectures, 31 sessions, and 22 posters. A variety of societies had virtual rooms at SPE 2021 including *Bernoulli Society*, *Brazilian Statistical Association*, *Caucus for Women in Statistics*, and the *International Society for Bayesian Analysis*—just to name a few. Institutional members of the Portuguese Statistical Society were also represented (e.g. Statistics Portugal, Banco de Portugal, PORDATA). For more details on the meeting please, see *www.spe2021.uevora.pt/en/*.

Recent Developments in Statistics and Data Science highlights some selected contributions that were presented at SPE 2021. This volume covers a broad range of topics lying at the interface between Statistics and Data Science, such as applied statistics, computational statistics, extremes and outliers, medical statistics, modeling time series and stochastic processes, and data visualization, among others.

And speaking of visualization, Fig. 1 depicts a word cloud of all the titles and abstracts in this volume. While this chart is not a substitute for a table of contents, it does not summarize the order by which the articles appear in this issue! it offers a visual roadmap of what is to be found ahead. Given the broad scope of topics covered, we have opted for clustering articles according to the similarity of topics.

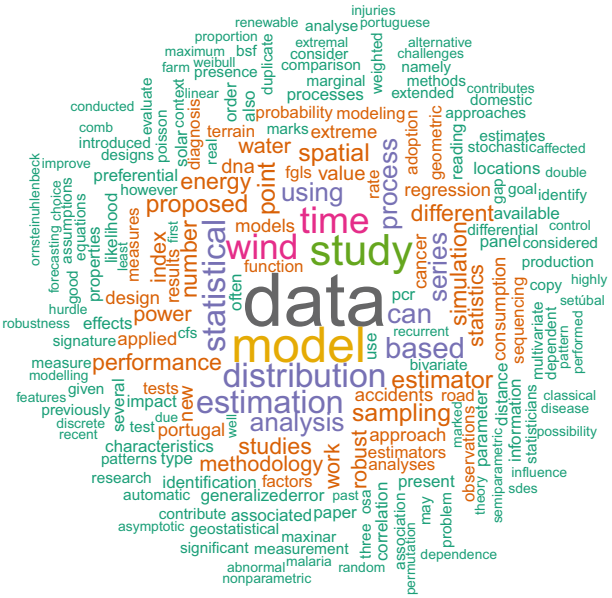

Fig. 1 Word cloud summarizing all titles, keywords, and abstracts of contributions included in this issue

And, as the tag cloud in Fig. 1 reveals, *data* are the common denominator across most contributions.

We are indebted to many people. First, we would like to thank the authors for their contributions and to everyone involved in the peer-review process who did a superb job on meeting tight deadlines in a thoughtful manner. They all worked hard so that the community keeps breaking new ground, and should be proud of their achievements. We are also indebted to Springer for their excellent collaboration on the production of this issue, and to the Scientific Committee and keynote speakers for their contributions to the meeting. Last but not the least, our words of thanks go to the organizers of SPE 2021 for their outstanding work, and to the community of the Portuguese Statistical Society for their continuous, and yet unbounded! support.

March 2022

Regina Bispo
Lígia Henriques-Rodrigues
Russell Alpizar-Jara
Miguel de Carvalho

Contents

How to Increase the Visibility of Statisticians in the Modern World of Dataism? 1
Nuno Sepúlveda

A Robust Hurdle Poisson Model in the Estimation of the Extremal Index 15
Manuela Souto de Miranda, M. Cristina Miranda, and M. Ivette Gomes

Computational Study of the Adaptive Estimation of the Extreme Value Index with Probability Weighted Moments 29
Frederico Caeiro and M. Ivette Gomes

Estimation of the Weibull Tail Coefficient Through the Power Mean-of-Order-p 41
Frederico Caeiro, M. Ivette Gomes, and Lígia Henriques-Rodrigues

On the Maximum of a Bivariate Max-INAR(1) Process 55
Sandra Dias and Maria da Graça Temido

The Performance of a Combined Distance Between Time Series 71
Margarida G. M. S. Cardoso and Ana Alexandra Martins

Zero-Distorted Generalized Geometric Distribution with Application to Time Series of Counts 85
Esmeralda Gonçalves and Diogo Sousa

Uncovering Abnormal Water Consumption Patterns for Sustainability's Sake: A Statistical Approach 99
Ana Borges, Clara Cordeiro, and M. Rosário Ramos

Modeling and Forecasting Wind Energy Production by Stochastic Differential Equations 109
Paulo Cabral and Paula Milheiro–Oliveira

Intensity-Dependent Point Processes 123
Andreia Monteiro, Maria Lucília Carvalho, Ivone Figueiredo, Paula Simões, and Isabel Natário

Geostatistical Sampling Designs Under Preferential Sampling for Black Scabbardfish 137
Paula Simões, Maria Lucília Carvalho, Ivone Figueiredo, Andreia Monteiro, and Isabel Natário

Modeling Residential Adoption of Solar Photovoltaic Systems 153
Carolina Goldstein, José Miguel Espinosa, and Regina Bispo

Comparison of Semiparametric Approaches to Two-Way ANOVA in the Presence of Heteroscedasticity 189
Dulce G. Pereira and Anabela Afonso

Some Determinants for Road Accidents Severity in the District of Setúbal 203
Paulo Infante, Anabela Afonso, Gonçalo Jacinto, Leonor Rego, Pedro Nogueira, Marcelo Silva, Vitor Nogueira, José Saias, Paulo Quaresma, Daniel Santos, Patrícia Gois, and Paulo Rebelo Manuel

Impact of Misclassification and Imperfect Serological Tests in Association Analyses of ME/CFS Applied to COVID-19 Data 215
João Malato, Luís Graça, and Nuno Sepúlveda

Identification of Antibody Responses Predictive of Protection Against Clinical Malaria 227
André Fonseca, Clara Cordeiro, and Nuno Sepúlveda

Statistical Challenges in Mutational Signature Analyses of Cancer Sequencing Data 241
Víctor Velasco-Pardo, Michail Papathomas, and Andy G. Lynch

PCR Duplicate Proportion Estimation and Consequences for DNA Copy Number Calculations 259
Andy G. Lynch, Mike L. Smith, Matthew D. Eldridge, and Simon Tavaré

A Retrospective Study on Obstructive Sleep Apnea 281
Ricardo São João, Andreia Cardoso, Tiago Dias Domingues, Marta Fradinho, Vânia Silva, and Amélia Feliciano

Censored Multivariate Linear Regression Model 293
Rodney Sousa, Isabel Pereira, and Maria Eduarda Silva

A Methodology to Reveal Terrain Effects from Wind Farm SCADA Data Using a Wind Signature Concept 309
Alda Carvalho, Daniel C. Vaz, Tiago A. N. Silva, and Cláudia Casaca

A Robust Version of the FGLS Estimator for Panel Data 325
Anabela Rocha and M. Cristina Miranda

The Extended Chen–Poisson Marginal Rate Model for Recurrent Gap Time Data 337
Ivo Sousa-Ferreira, Cristina Rocha, and Ana Maria Abreu

On Classical Measurement Error within a Bayesian Nonparametric Framework 353
Emmanuel Bernieri and Miguel de Carvalho

Author Index 363