

Statistics, Technology and Data Science for Economic
and Social Development

Book of short papers of the ASA Rome Conference
18 to 20 September 2024

“Measuring and Interpreting World Changes with Statistics,
Data Science and AI”

Book of Short Papers of the
Sapienza University of Rome Conference

Supplement to Volume 37/2

Guest Editors:

Leonardo S. Alaimo, Fabio Crescenzi,
Luigi Fabbris, Filomena Maggino, Maurizio Vichi

Statistica Applicata - ITALIAN JOURNAL OF APPLIED STATISTICS

EDITORIAL TEAM

EDITOR IN CHIEF

- Francesco Palumbo, Università di Napoli Federico II, Naples, Italy

CO-EDITORS ON A SPECIFIC SUBJECT

- Alessandro Celegato, AICQ Centronord - Quality and technology in production
- Adriano Decarli, Università di Milano, IRCCS /INT Foundation, Milan, Italy - Social and health studies
- Luigi Fabbris, Università di Padova, Padua, Italy - Surveys and experiments
- Vittorio Frosini, Università Cattolica del Sacro Cuore, Milan, Italy - Book review
- Antonio Giusti, Università di Firenze, Florence, Italy - Data Science
- Paolo Mariani, Università di Milano Bicocca, Milan, Italy - Social and economic analysis and forecasting

SCIENTIFIC COMMITTEE

- Thomas Aluja, UPC, Barcelona, Spain
- Paul P. Biemer, RTI and IRSS, Chicago, USA
- Jörg Blasius, Universität Bonn, Bonn, Germany
- Irene D'Epifanio, Universitat Jaume I, Castelló de la Plana, Spain
- Vincenzo Esposito Vinzi, ESSEC Paris, France
- Gabriella Grassia, Università di Napoli Federico II, Naples, Italy
- Michael J. Greenacre, UPF, Barcelona, Spain
- Salvatore Ingrassia, Università di Catania, Catania, Italy
- Ron S. Kenett, KPA Ltd. and Samuel Neaman Institute, Technion, Haifa, Israel
- Stefania Mignani, Università di Bologna Alma Mater, Bologna, Italy
- Tormod Naes, NOFIMA, Oslo, Norway
- Alessandra Petrucci, Università di Firenze, Florence, Italy
- Monica Pratesi, Università di Pisa, Pisa, Italy
- Maurizio Vichi, Sapienza Università di Roma, Rome, Italy
- Giorgio Vittadini, Università di Milano Bicocca, Milan, Italy
- Adalbert Wilhelm, Jacob University, Breimen, Germany

ASSOCIATE EDITORS

- Francesca Bassi, Università di Padova, Padua, Italy
- Bruno Bertaccini, Università di Firenze, Florence, Italy
- Matilde Bini, Università Europea, Rome, Italy
- Giovanna Boccuzzo, Università di Padova, Padua, Italy
- Maurizio Carpita, Università di Brescia, Brescia, Italy
- Giuliana Coccia, ISTAT, Rome, Italy
- Fabio Crescenzi, ISTAT, Rome, Italy
- Franca Crippa, Università di Milano Bicocca, Milan, Italy
- Corrado Crocetta, Università di Foggia, Foggia, Italy

- Cristina Davino, Università di Napoli Federico II, Naples, Italy
- Loretta Degan, Gruppo Galgano, Milan, Italy
- Tonio Di Battista, Università di Chieti-Pescara “Gabriele D’Annunzio”, Pescara, Italy
- Tommaso Di Fonzo, Università di Padova, Padua, Italy
- Francesca Di Iorio, Università di Napoli Federico II, Naples, Italy
- Simone Di Zio, Università di Chieti-Pescara “Gabriele D’Annunzio”, Pescara, Italy
- Filippo Domma, Università della Calabria, Rende, Italy
- Alessandra Durio, Università di Torino, Turin, Italy
- Monica Ferraroni, Università di Milano, Milan, Italy
- Giuseppe Giordano, Università di Salerno, Salerno, Italy
- Michela Gnaldi, Università di Perugia, Perugia, Italy
- Domenica Fioredistella Iezzi, Università di Roma Tor Vergata, Rome, Italy
- Michele Lalla, Università di Modena e Reggio Emilia, Modena, Italy
- Maria Cristina Martini, Università di Modena e Reggio Emilia, Modena, Italy
- Fulvia Mecatti, Università di Milano Bicocca, Milan, Italy
- Sonia Migliorati, Università di Milano Bicocca, Milan, Italy
- Michelangelo Misuraca, Università della Calabria, Rende, Italy
- Francesco Mola, Università di Cagliari, Cagliari, Italy
- Roberto Monducci, ISTAT, Rome, Italy
- Isabella Morlini, Università di Modena e Reggio Emilia, Modena, Italy
- Biagio Palumbo, Università di Napoli Federico II, Naples, Italy
- Alfonso Piscitelli, Università di Napoli Federico II, Naples, Italy
- Antonio Punzo, Università di Catania, Catania, Italy
- Silvia Salini, Università di Milano, Milan, Italy
- Luigi Salmaso, Università di Padova, Padua, Italy
- Germana Scepi, Università di Napoli Federico II, Naples, Italy
- Giorgio Tassinari, Università di Bologna Alma Mater, Bologna, Italy
- Ernesto Toma, Università di Bari, Bari, Italy
- Rosanna Verde, Università della Campania “Luigi Vanvitelli”, Caserta, Italy
- Grazia Vicario, Politecnico di Torino, Turin, Italy
- Maria Prosperina Vitale, Università di Salerno, Salerno, Italy
- Susanna Zaccarin, Università di Trieste, Trieste, Italy
- Emma Zavarrone, IULM Milano, Milan, Italy

EDITORIAL MANAGER

- Domenico Vistocco, Università di Napoli Federico II, Naples, Italy

EDITORIAL STAFF

- Antonio Balzanella, Università della Campania “Luigi Vanvitelli”, Caserta, Italy
- Luca Bagnato, Università Cattolica del Sacro Cuore, Milan, Italy
- Paolo Berta, Università di Milano Bicocca, Milan, Italy
- Francesca Giambona, Università di Firenze, Florence, Italy
- Rosaria Romano, Università di Napoli Federico II, Naples, Italy
- Rosaria Simone, Università di Napoli Federico II, Naples, Italy
- Maria Spano, Università di Napoli Federico II, Naples, Italy

A.S.A CONTACTS**Principal Contact**

Francesco Palumbo (Editor in Chief)

editor@sa-ijas.org

Support Contact

Domenico Vistocco (Editorial Manager)

ijas@sa-ijas.org

JOURNAL WEBPAGE<https://www.sa-ijas.org/ojs/index.php/sa-ijas>

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (<http://web.it.nctu.edu/~qtqm/>)

SINERGIE – Italian Journal of Management



Statistica Applicata – Italian Journal of Applied Statistics (ISSN:1125-1964, E-ISSN:2038-5587) applies the Creative Commons Attribution (CC BY) license to everything we publish.

Published: April 2025

© 2025 Author(s)

Content license: except where otherwise noted, the present work is released under Creative Commons Attribution 4.0 International license (CC BY 4.0: <https://creativecommons.org/licenses/by/4.0/legalcode>). This license allows you to share any part of the work by any means and format, modify it for any purpose, including commercial, as long as appropriate credit is given to the author, any changes made to the work are indicated and a URL link is provided to the license.

Metadata license: all the metadata are released under the Public Domain Dedication license (CC0 1.0 Universal: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>).

Published by Firenze University Press

Powered by Firenze University Press

Firenze University Press

Università degli Studi di Firenze

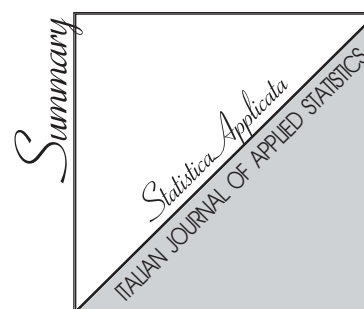
via Cittadella, 7, 50144 Firenze, Italy

www.fupress.com

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising: CLEUP sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.

Rules for manuscript submission: <https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions>

Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA, 7 March 2023.



Introduction	13
Determinants of cooperation among hotels in Venice: A network approach	15
<i>Amati, V., Moretti, A., Finzi, R.M.</i>	
Modelling default risk in Italian SMEs: A spatial probit approach	21
<i>Amoruso, E., Bruno, E., Castellano, R., Punzo, R.</i>	
The influence of UNESCO sites on tourism demand and supply: Insights from Italian regions .	25
<i>Antolini, F., Cesarini, S., Terragli, I.</i>	
“Inhumane”: Quantitative textual analysis of a qualitative survey on gender violence	31
among victims of international human trafficking	
<i>Arcaio, M., Barbaro, F., Parroco, A.M.</i>	
Keyword-enhanced summarization from scientific literature through an integrated approach .	37
<i>Aria, M., Cuccurullo, M., D’Aniello, L., Misuraca, M., Spano, M.</i>	
A composite indicator to measure attitude towards sustainable tourism by European citizens ..	43
<i>Bassi, F., Scioni, M.</i>	
Data quality assessment framework for composite indicator construction: An application to service delivery	49
<i>Bassi, F., Muhanguzi, H., Scioni, M.</i>	
Exploring art market trends: A Delphi study on collectible design	55
<i>Benedan, L., Codignola, F., Galeone, C., Mariani, P.</i>	
Resilience versus depression of Italian youth after the COVID-19 pandemic	61
<i>Bollani, L., Cascella, C., Fabbris, L.</i>	
Trust in health science and organizations in Italy during the Covid-19 pandemic	67
<i>Bollani, L., Fabbris, L.</i>	
Artificial intelligence and future scenarios: Preliminary results from an application on family policies in north-eastern Italy	73
<i>Bolzan, M., Calleo, Y., Di Zio, S., Marozzi, M., Scioni, M.</i>	

A non-parametric test for comparative evaluations of students' performances	79
<i>Bonnini, S., Borghesi, M., Giacalone, M., Piscopo, G.</i>	
An empirical analysis of various methods for making inferences with non-probability Samples	85
<i>Braitto, L., Rocco, E.</i>	
A statistical approach for the predictive analysis of insolvency in Italy	91
<i>Brentari, E., Ermici, G., Zanetti, A.</i>	
An innovative point of view for statistical dissemination to prevent road accidents using the 'collision matrix'	97
<i>Broccoli, M., Bruzzone, S.</i>	
A new algorithm for the correction and control of road accidents data based on string similarities.....	103
<i>Bruzzone S., Valentino, L., Moretti D., Tavanti M., De Blasi, F.</i>	
Advancing spatial scenarios adopting generative pre-trained transformers and text-to-image models.....	109
<i>Calleo, Y., Pilla, F., Di Zio, F.</i>	
Enhancing mobility and tourism through data analytics and generative AI	115
<i>Cappellari, D., Mantegar, G., Zamengo, B.</i>	
An innovative approach for the analysis of socio-economic phenomena at sub-municipal level: The Household Deprivation Study project.....	121
<i>Carbonetti, G., Biasciucci, F., Cutillo, A., Mazziotta, M., Quondamstefano, V., Tronu, D.</i>	
Volatility decomposition of traffic flow time series with complex seasonality using GARCH models.....	127
<i>Carpita, M., Metulini, R., Migliorati, M.</i>	
Economic well-being of households and attractiveness of territories: The case of Italian large municipalities.....	133
<i>Casacci, S.</i>	
An application of the partial credit model to measure the resilience of Italian young adults in the aftermath of the COVID-19 pandemic	139
<i>Cascella, C., Bollani, L., Fabbris, L.</i>	
Data science approaches to the study of multimorbidity patterns at older ages.....	145
<i>Caselli, N., Dang, L., H., K., Rettaroli L., Miglio, R.</i>	
Analysing fan engagement and sentiment on social media following Bologna FC's 2023/2024 Champions League qualification	151
<i>Cavrini, G., Stavolo, A.</i>	
Text mining of tourist reviews: Sentiment analysis and trends in the Apulian tourism.....	157
<i>Centorame, D., Pino, G., Sarra A., Nissi, E.</i>	
An exploratory analysis of the comments from the Istat's household budget survey questionnaires for designing an App	163
<i>Cezza, E., D'Ambrosio, G., De Gaetano, L., Raimondi, E.</i>	

Combining expert knowledge and data-driven classification to support prediction and validation of census population counts.....	169
<i>Chieppa, A., Bernardini, A., Cibella, N., Solari, N.</i>	
The “group approach” versus the “inverted cascade approach” for calculating the super-dividend test of holding entities	175
<i>D’Achille, A., Boscaino, F.,</i>	
Migration as a diffusion process: The case of Bologna municipality.....	181
<i>D’Attoma, I., Cid, C., Tassinari, C.</i>	
Users beyond “boundaries”: Recent experiences of promoting statistical literacy by Istat at the territorial level among non-regular data users	187
<i>Da Valle, S., Osti, S., Viviano, L.C.M.</i>	
Conceptualizing and measuring educational poverty through family, school, and environmental opportunities.....	193
<i>Davino, C., De Falco, A., Fabbriatore, R., Gherghi, M., Romano, R.</i>	
The Italian NEETs after the COVID-19: What should we expect for the future? Different approaches to forecasts	199
<i>De Luca, G., Rocca, A., Mazzocchi, P., Quintano C.</i>	
AI and geostatistics: Innovative solutions for reducing and storing CO2	205
<i>Digrandi, A., M., Cimmino, P.</i>	
Leveraging Industry 5.0 for official business statistics innovation	211
<i>Distefano, D., Bosso, P., Di Paolo, G., G., Papa, P.</i>	
Future orientation and psychological and social capital of Italian young adults	217
<i>D’Uggento, A.M., Fabbris, L.</i>	
Travel and vacations of Italian senior tourists	223
<i>D’Uggento, A.M., d’Ovidio, F., D, Marin, C., Pepe, I., Toma, E.</i>	
It’s how you use the items that counts: An intelligent procedure for item selection in Item response theory.....	229
<i>Epifania, O., M., Anselmi, P., Robusto, E.</i>	
School socioeconomic status and students’ achievements: Exploring territorial heterogeneity through a multilevel regression discontinuity approach	235
<i>Fabbriatore, R., Sannino, P., Davino, C.</i>	
Capacity to aspire in high school students: Theoretical definition and measurement	241
<i>Fabbriatore, R., Romano, R., Davino, C.</i>	
Improving online misogyny detection: The role of human annotation and explanation guided learning	247
<i>Fontanella, L., Cucco, A., del Gobbo, E., Ignazzi, E., Fontanella, S.</i>	
The evolution of Explainable Artificial Intelligence (XAI): A preliminary systematic literature view.....	253
<i>Gnasso, A., Aria, M.</i>	

Innovative ER solutions: Using data to drive operational success.....	259
<i>Grassia, M.G., Sacco, D., Paesano, D., Dagostino, F., Annachiarico, A., Di Palma, V.</i>	
Responsible AI/ML implementation in official statistics: A framework for decision-making.....	265
<i>Grippo, G., Righi, A.</i>	
Copula functions and dynamic hydrological data: The case of Scotland	269
<i>Guerini, S.S., Spezia, L.</i>	
AI-generated sentiment in state-space models for Toyota Camry US volume sales forecasting	275
<i>Iezzi, D., F., Monte, R., Pasquini, D.</i>	
Modeling virality in social media.....	281
<i>Iezzi, D., F., Monte, R., Pasquini, D.</i>	
Organizational flexibility measures: Remote working at the Italian National Institute of Statistics	287
<i>Limiti, C., Sola G.</i>	
Improving information on labour demand for decision-making by integrating traditional surveys, administrative sources, and web-based data	293
<i>Lucarelli, A., Righi, A., Rizzi R.</i>	
Early detection of university dropouts: An experiment using video-lectures access data	299
<i>Martini, M.C.</i>	
Statistical indicators for the analysis of social sustainability and human well-being.....	305
<i>Massari, A., Perchinunno, P., L'Abbate, S., Leogrande, D., Mongelli, D.</i>	
Health and discrimination among refugees in Germany	311
<i>Mendola, D., Busetta, A.</i>	
Estimating overdispersion in non-transformed scores of ANVUR's department ranking procedure: A Beto-multinomial model.....	317
<i>Montanari, G., E., Doretti, M.</i>	
Design of a test for assessing teamwork and collaborative skills with competence-based knowledge space theory	323
<i>Morleo, F., Vitanza, A., Anselmi, P.</i>	
Efficiency of Italian courts: Awareness from spatial stochastic frontier analysis	329
<i>Nissi, E., Sarra, A., Evangelista, A., Di Battista, T.</i>	
Campanian healthcare facilities under the microscope: Reviews, polarity, and georeferencing	335
<i>Paesano, S., Grassia, M.G., Marino, M., Mazza, R., Sacco, D., Stavolo, A.</i>	
Markov switching autoregressive models and Bayesian hydrology	341
<i>Paroli, R., Spezia, L.</i>	
Collecting ADS-B data for timely production of air traffic statistics	347
<i>Pietropaoli, S., Amarone, M.</i>	

A statistical approach to measuring the impact of medication use on the lived experience of patients with inflammatory bowel disease	353
<i>Piscitelli, A., D'Uggento, A., M., Equatore, E., Principi, M.B.</i>	
Artificial intelligence in support of accessibility: The Brain-friend device	359
<i>Santarcangelo, V., Crisafulli, S.G., Giacalone, M., Piscopo, G., Massa, E., Vitullo, S., Scintu, D., Romano, A., Radosti, G.</i>	
Combining textual analysis and IRT to interpret the students' voice.....	365
<i>Sarra, A., Fontanella, L., Evangelista, A., Di Battista, T.</i>	
Exploring teachers' professional identity: A case study utilizing structural equation modeling and text mining in a multimodal approach	371
<i>Sarra, A., Pentucci, M., Fontanella, L.</i>	
Selecting sample households with high quality addresses for population census: Problems for field survey network, but bias in the estimates?.....	377
<i>Scavalli, E.</i>	
Statistical analysis of smartphone mobility data for air quality assessment.....	383
<i>Shaukat, M., H., Finazzi, F.</i>	
The combination of expert knowledge and Bayesian networks to identify the relevant factors in diabetic kidney disease (DKD).....	389
<i>Slanzi, D., Poli, I., Jones, R., D., Mayer, G.</i>	
Understanding consumer behavior in the plant-based meat market: Insights from Bayesian network analysis.....	395
<i>Slanzi, D., Lanzini, P., Checchinato, F., Colapinto, C.</i>	
Exploring childfree and childless on social media: A text analysis approach	401
<i>Stavolo, A., Cavrini, G., Cisotto, E., Marino, M.</i>	
Mass tourism and local housing: An exploratory textual analysis.....	407
<i>Stavolo, A., Cataldo, R., Grassia, M., G., Marino, M.</i>	
Multiverse analysis: Heuristic for multi-model inference or theory of the model error?	413
<i>Tomaselli, V., Cantone, G., G.</i>	
From crisis to crisis: Mapping material deprivation in the EU through the 2000s.....	419
<i>Traversa, S., Ivaldi, E.</i>	
The effect of user web journey on browsing behaviour and buying attitude: The case of Sunweb.nl.....	425
<i>Urso, F., Argentino, N., Abbruzzo, A., Mohammadi, R., Pak, K., Cracolici, M., F.</i>	
Geography of business alliances and spatial network complexity: The backbone of formal collaboration agreements	429
<i>Vurro, A., E., Bumba, A., Giuffrida, A., Mazzitelli, A., Espa, G.</i>	

Spatial analysis of climate change awareness across countries	435
<i>Zammarchi, G., Maranzano, P.</i>	
Future scenarios: A financial network analysis of stock prices in emerging metaverse market.....	441
<i>Zavarrone, E., Forciniti, A.</i>	

Introduction

This volume brings together a selection of papers presented at the scientific conference of the ASA (Association for Applied Statistics), held in Rome from 18 to 20 September 2024. It retains the title of the conference, “Measuring and Interpreting World Changes with Statistics, Data Science and AI.” The majority of papers focuses on modeling the relationships between statistics and the emerging disciplines of data science and artificial intelligence (AI), while others address areas of applied statistics and methodological advances in official statistics.

At a time when global challenges—such as climate change, economic inequality, and technological transformation—are reshaping societies, the role of statistics and data science in interpreting and guiding these changes is more crucial than ever. This volume aims to offer a comprehensive perspective on how statistical reasoning, enhanced by modern tools from AI and data science, can contribute to understanding and addressing complex real-world issues.

All papers were anonymously reviewed by at least two independent experts from the conference’s scientific committee. The published articles reflect a rigorous peer-review process and incorporate suggestions and revisions—both substantive and formal—made by the authors in response to expert feedback.

Conference presentations were grouped into thematically homogeneous sessions, covering topics such as official statistics, urban analytics, labor market dynamics, and innovations in data integration. Compared to the oral presentations, the papers in this volume have been expanded and refined, offering readers a more in-depth exploration of the subjects discussed.

Due to the self-selection of contributors and the rigorous evaluation process, the number of eligible papers has been significantly reduced. As a result, the content-based session structure could not be preserved in this publication. Instead, the articles are presented in alphabetical order by surname of the first author.

The guest editors would like to express their sincere thanks to the organizing committee and all reviewers and authors for their invaluable contributions to the success of the conference and this volume. We believe this collection will be of interest to researchers, practitioners, and policymakers engaged in quantitative analysis and evidence-based decision-making. We are confident that the insights shared in these contributions will foster further research and debate in the fields of applied statistics, data science, and artificial intelligence.

This volume will be published as a supplement to a regular issue of the international journal *Statistica Applicata – Italian Journal of Applied Statistics*. All articles are freely available in open access format. When citing a paper, please use the following format:

Author(s) (Year). Paper title. In L.S. Alaimo, F. Crescenzi, L. Fabbri, F. Maggino (eds.), *Measuring and Interpreting World Changes with Statistics, Data Science and AI. Book of Short Papers of the ASA Rome Conference* (pages). Padua: Cleup.

The guest-editors:

Leonardo S. Alaimo is tenure track researcher in Social statistics at the Department of Statistical Sciences, Sapienza University of Rome

Fabio Crescenzi is a statistical advisor and former senior researcher at Istat (Italian National Institute of Statistics), Rome

Luigi Fabbri is president of the Association for Applied Statistics and former professor of Social statistics at the University of Padua, Italy.

Filomena Maggino is full professor of Social statistics at the Department of Statistical Sciences, Sapienza University of Rome

Maurizio Vichi is a former professor of Statistics at the Department of Statistical Sciences, Sapienza University of Rome

Other members of the Scientific committee:

Fabrizio Antolini (University of Reramo)
Margaret Antonicelli (Sapienza University of Rome)
Francesca Bassi (University of Padua)
Rossella Berni (University of Florence)
Bruno Bertaccini (University of Florence)
Matilde Bini (European University of Rome)
Giovanna Boccuzzo (University of Padua)
Luigi Bollani (University of Turin)
Eugenio Brentari (University of Brescia)
Maurizio Carpita (University of Brescia)
Giulia Cavrini (University of Bolzano-Bozen)
Alessandro Celegato (AICQ-AISS, PSV Project Service and Value)
Giuliana Coccia (ASViS)
Cristina Davino (Federico II University of Naples)
Adriano Decarli (University of Milan)
Loretta Degan (Galvano Group Co., Milan)
Enrico di Bella (University of Genoa)
Angela Maria Digrandi (CNR-National Council for Research)
Simone Di Zio (G. D'Annunzio University of Chieti and Pescara)
Viviana Egidi (formerly Sapienza University of Rome)
Benito Vittorio Frosini (Sacred Heart Catholic University of Milan)
Antonio Giusti (University of Florence)
M. Gabriella Grassia (Federico II University of Naples)
Andrea Guizzardi (University of Bologna)
D. Fioredistella Iezzi (Tor Vergata University of Rome)
Paolo Mariani (University of Milan-Bicocca)
Stefania Mignani (University of Bologna)
Francesco Palumbo (Federico II University of Naples)
Chiara Parretti (Marconi University)
Alfonso Piscitelli (Federico II University of Naples)
Tommaso Pronunzio (Assirm, Ales Research)
Alessandra Righi (Istat, Rome)
Rosaria Romano (Federico II University of Naples)
Vito Santarcangelo (iInformativa, Matera)
Giuseppe Sindoni (Istat, Rome)
Giorgio Tassinari (University of Bologna, co-chair)
Nicola Tedesco (University of Cagliari)
Venera Tomaselli (University of Catania)
Maurizio Vichi (co-chair, La Sapienza University of Rome)
Domenico Vistocco (Federico II University of Naples)
Giorgio Vittadini (University of Milan-Bicocca)
Mariangela Zenga (University of Milan-Bicocca)

Spatial analysis of climate change awareness across countries

Gianpaolo Zammarchi^a, Paolo Maranzano^b

^a Department of Economics and Business Science, University of Cagliari, Cagliari, Italy.

^b Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Milan, Italy.

1. Introduction

One of the greatest challenges that current and future generations will face is climate change and the extreme or severe repercussions this will entail. For instance, the increasingly frequent extreme weather events, the melting polar ice caps and warming oceans (with all the repercussions that these can have on the life of the animal species that live, hunt or base their life cycle on these places), and the rising sea levels are all effects of climate change. All countries are impacted by climate change, however certain areas may be more susceptible to particular effects. According to the Germanwatch observatory's annual assessment of the global climate risk index 2021, developing nations are more susceptible to the effects of climate change due to their reduced ability to adapt (Global Climate Risk Index, 2021). In fact, low- to lower-middle income countries account for eight of the ten most severely affected countries in terms of fatalities and economic losses caused by extreme weather events (such as storms, floods, heatwaves, etc.) in 2019. Three of the most affected countries were Mozambique, Zimbabwe and the Bahamas.

Increasing public awareness of the causes and effects of climate change is crucial because it can motivate policymakers to take action to cut greenhouse gas emissions and encourage individual behavioural adjustments. Since there are variations in how different nations might be vulnerable to some of the effects of climate change, the goal of this work was to examine how different nations differ in terms of their degree of knowledge of the problem.

Furthermore, we investigated the connection between awareness and pertinent economic indicators like GDP, Gini index, spending on environmental preservation, as these variables represent relevant characteristics of countries and present the preliminary results of an analysis based on a hierarchical clustering using spatial constraints.

2. Methodology

Two editions of the “International public opinion on climate change” survey were used to gather awareness data. This survey's main goals were to examine people's knowledge, beliefs, attitudes, policy preferences, and behaviour related to climate change. We collected data pertaining to the “climate awareness” question. In addition, we gathered data on GDP (GDP by Country, 2022) and environmental protection expenditures (Environmental Protection Expenditures, 2023), for the nations that were part of the study. Moreover, we created an extended version of the dataset with data collected from the Climate Change Knowledge Portal, the Penn World Table, the Global Carbon Atlas, the Standardized World Income Inequality Database, and the Our World in Data web portal, in order to be able to apply the spatially clustered methodology. Hierarchical clustering is a cluster analysis method that aims to create a hierarchy between clusters. There are two alternative approaches, that is, “bottom-up” (agglomerative) or “top-down” (divisive) and in both approaches

the concept of distance and its choice are of fundamental importance. The results of hierarchical clustering are usually shown in a graph called a dendrogram. Partitioning clustering, like K-means for instance, aims at dividing the data set into a specific number of clusters to minimize within-cluster variance while maximizing between-cluster variance. The algorithm assigns iteratively each of the n observations to one of the k clusters based on the distance between the observation and the cluster center. The main issue with K-means is that the number k of clusters is a hyperparameter that must be provided by the user.

In the first part of the analysis, we used data from the 2022 edition of the survey, including responses from 108,946 Facebook users located in 110 areas worldwide (Leiserowitz et al., 2022) as well as the 2021 edition of the survey, including responses from 76,328 users located in 30 countries (Leiserowitz et al., 2021). To investigate variations in climate change awareness across time, we incorporated information from the 30 nations that were included in both survey editions. Using the four response items, a pairwise Euclidean distance matrix was calculated in order to group nations based on their levels of awareness. As shown in Equation 1, single observations were combined into clusters using the Ward's approach.

$$\frac{|A| \cdot |B|}{|A \cup B|} \|\mu_A - \mu_B\|^2 = \sum_{x \in A \cup B} \|x - \mu_{A \cup B}\|^2 - \sum_{x \in A} \|x - \mu_A\|^2 - \sum_{x \in B} \|x - \mu_B\|^2 \quad (1)$$

where A and B are two sets of observations.

Based on the elbow method, an optimal number of clusters was identified for both data sets. We created a hierarchical cluster using the `hclust` function in the `stats` R package (R Core Team, v. 4.3). The `dendextend` R package (Galili, 2015) was used to compare the two hierarchical clusters in order to highlight cluster variations that have occurred over the course of the two years. By matching the labels on the left side of the “double” hierarchical plot to the labels on the right, we were also able to calculate the “entanglement” value. The procedure is straightforward. Let v_1 and v_2 be two vectors corresponding to the order from 1 to n in which the labels appear on the left and right side of the plot, respectively. Moreover, let $\|\cdot\|_L$ be the L-norm distance between these two vectors. In order to obtain a maximum value M for this distance, the distance for the worst-case scenario is computed (which is when one side is the exact opposite of the other side). Finally, the “entanglement” can be written as: $E = \frac{\|v_1 - v_2\|_L}{M}$, with $E \in [0, 1]$. The two extreme scenarios, where $E = 0$ and $E = 1$, denote a perfect match between the two sets of labels and the scenario where the right tree is the left tree's full opposite, respectively. In addition, we applied K-means clustering to data from the 2022 edition of the survey with the goal of identifying groups of nations based on awareness, environmental protection expenditures, GDP, and knowledge of climate change. For this analysis, after merging and aligning data from different sources, we were able to use data from 62 countries.

The second part of the analysis is aimed at performing a hierarchical clustering with geographical constraints using an extended version of the dataset. In particular, we collected data from the above-mentioned sources to use the spatial clustering methodology proposed by Chavent et al. (2018). This method exploits two dissimilarity matrices called D_0 and D_1 along with a mixing parameter α that regulates the importance of the two matrices. The D_0 matrix provides information about dissimilarity in the “feature space”, while the D_1 matrix in the “constraint space”. Let I_α be the inertia, then the mixed pseudo inertia of cluster C_k^α can be defined as in Equation 2

$$I_\alpha(C_k^\alpha) = (1 - \alpha) \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{0,ij}^2 + \alpha \sum_{i \in C_k^\alpha} \sum_{j \in C_k^\alpha} \frac{w_i w_j}{2\mu_k^\alpha} d_{1,ij}^2 \quad (2)$$

where w_i is the weight of the i -th observation for $i = 1, \dots, n$ which represents the importance of

observation i (similarly w_j denotes the importance of observation j), $\mu_k^\alpha = \sum_{i \in C_k^\alpha} w_i$ is the weight of C_k^α , $d_{0,ij}^2$ and $d_{1,ij}^2$ are the $n \times n$ normalized dissimilarity matrices between observations i and j in D_0 and D_1 , respectively. Basically, the procedure performs a clustering based on the features collected to describe each country, while applying constraints (in this case spatial constraints) to create a combination that balanced the two aspects as a function of the value of α .

3. Results

In the first part of the analysis, we compared the two hierarchical clustering plots (2021 and 2022) to identify potential variations in the countries that were part of the clusters (Figure 1). The comparison of the two hierarchical clusters yielded an entanglement value of 0.29, indicating the existence of a discernible variation in the membership between the two years. Thailand and Colombia relocated to Cluster 3 from Cluster 2, i.e. in a cluster with poorer awareness of climate change, and Spain also experienced a decline in awareness, going from the cluster with the highest overall awareness to the intermediate one.

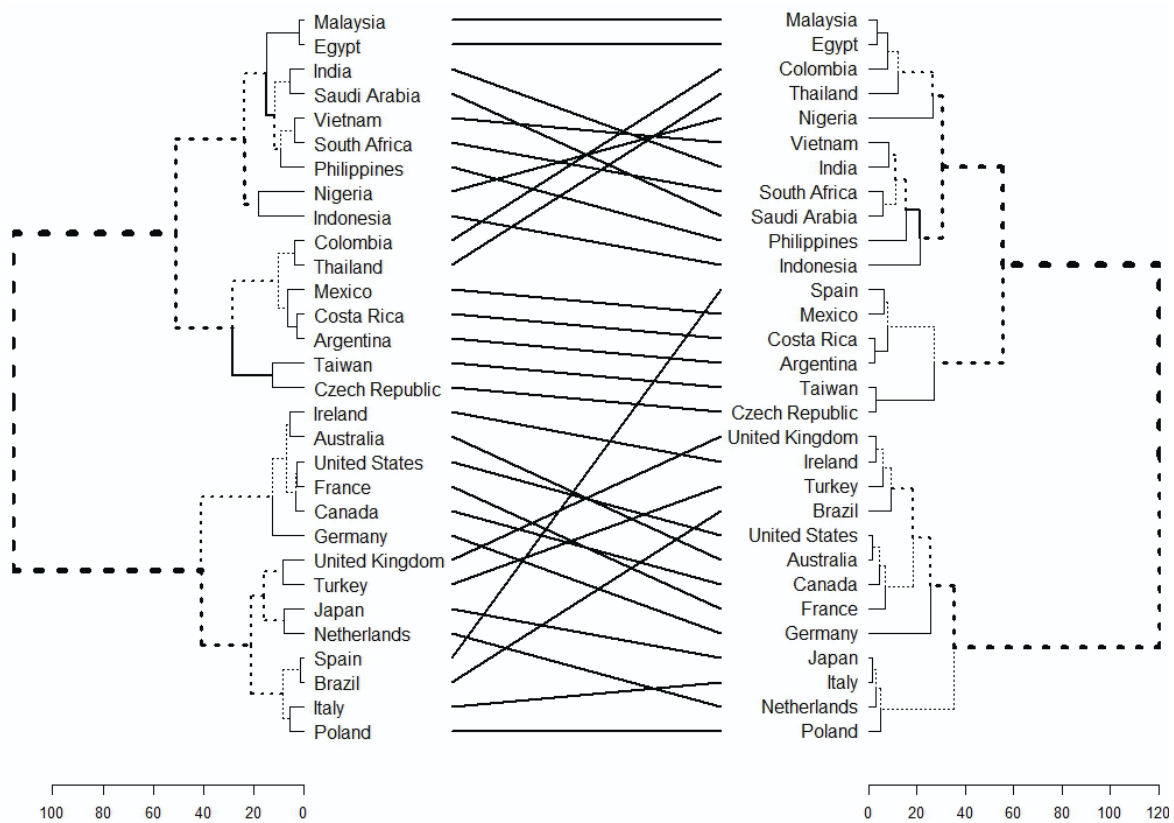


Figure 1. Comparison between awareness measured in 2021 and 2022

Using K-means, three clusters were identified based on the GDP, the environmental protection spending, and the levels of awareness: 1) countries with high awareness, GDP, and expenditure (green); 2) countries with low awareness, GDP, and expenditure (red); 3) countries with high GDP and awareness, but medium expenditure (Figure 2).

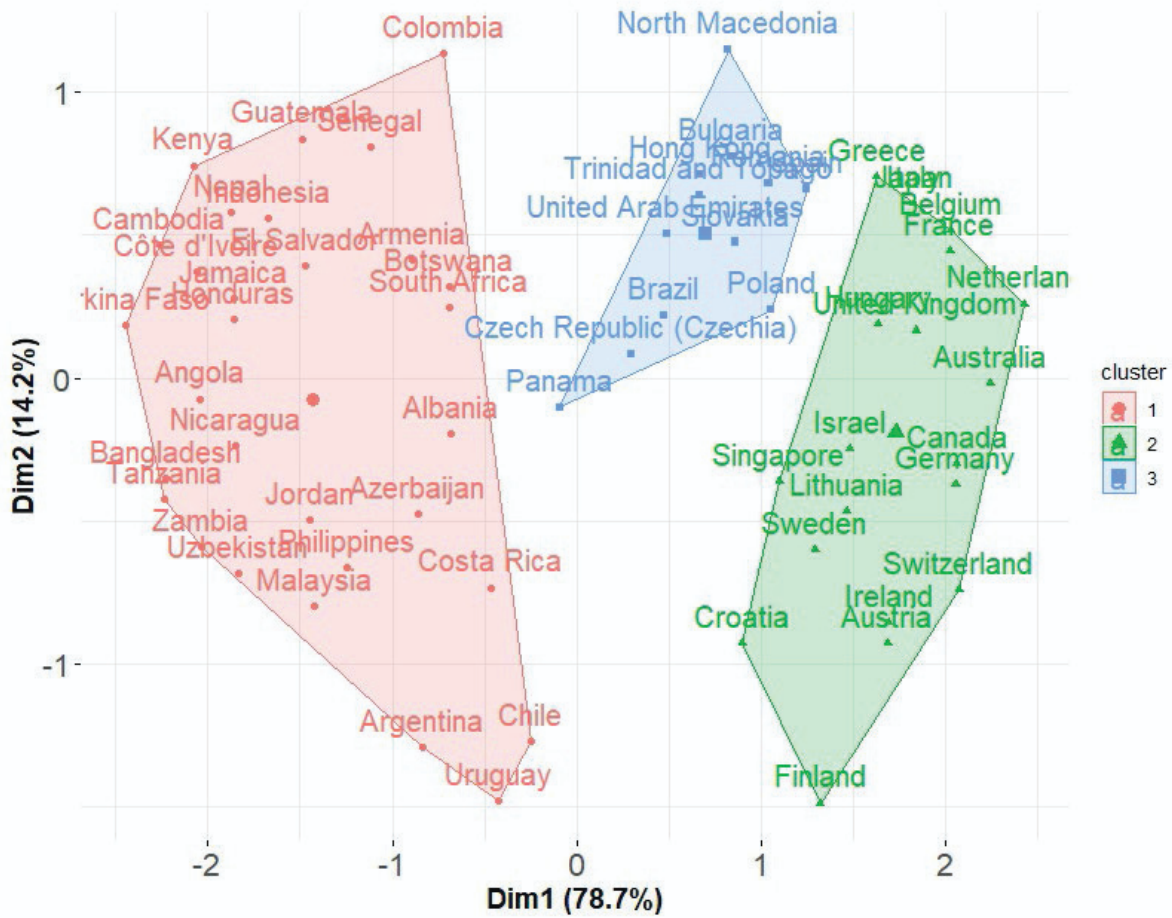


Figure 2. Clusters based on GDP, environmental protection spending, and awareness

Lastly, in Figure 3 we present a world map for all the countries analysed where the clustering was performed using the two matrices of features (based on awareness about climate change) and geographical dissimilarity. We selected an $\alpha = 0.55$ as a trade-off between D_0 and D_1 and set the number of clusters equal to 5. This figure shows that the geographical component is predominant, but nevertheless in some cases countries were not included in the most expected geographical cluster, but in another (e.g. Spain was not included in the European cluster but in the African one). Furthermore, Figure 4 provides additional information on the mean awareness levels of the countries in each cluster.

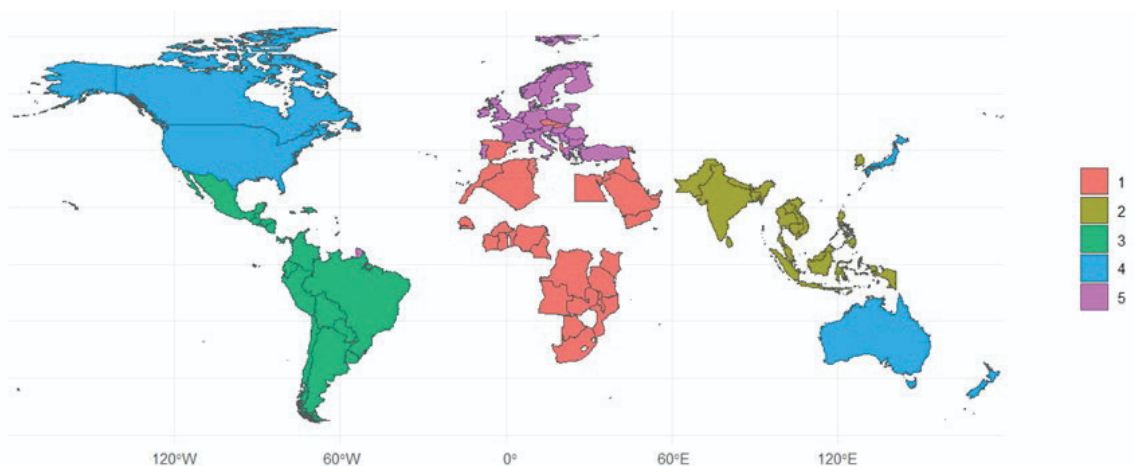


Figure 3. Clusters based on D_0 and D_1 matrices of features and geographical dissimilarities

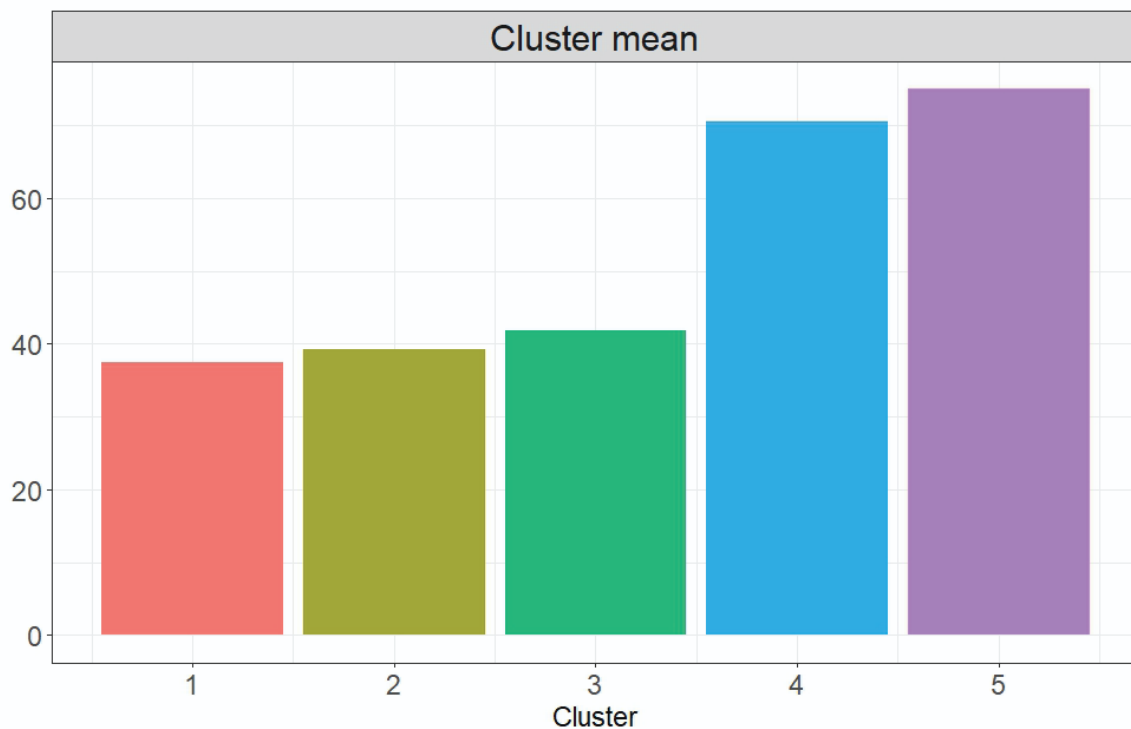


Figure 4. Barplot of awareness levels for the clusters created combining the D_0 and D_1 matrices

4. Conclusion

In this preliminary analysis, we studied the relationship between public awareness of climate change and variables such as GDP and environmental protection spending for each country considered. We performed a cluster analysis of nations according to their awareness of climate change and created three groups using economics and environmental variables. Finally, we produced an extended version of our dataset to be able to find the best value for α and then we used that value to create clusters based on both the features dissimilarity matrix and the geographical dissimilarity matrix. The clustering results consistently demonstrate that countries with high awareness levels are the richest countries (mostly Europeans and north Americans), while countries with low awareness are south Americans, Africans and part of the Asian. These countries could focus on educational programs and might share resources and plan strategies together to better address the problems related to increasing climate change awareness levels.

Future improvements of this work will include to expand the analysis and to explore the spatial correlations between countries in order to assess whether geographically neighbouring nations share a comparable level of awareness on climate change.

References

- Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J. (2018). ClustGeo: An R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4), pp. 1799-1822.
- Climate Change Knowledge Portal (2024). Last access: July 31, 2024, <https://climateknowledgeportal.worldbank.org/>
- Environmental Protection Expenditures (2023). Last access: December 15, 2023, <https://climatedata.imf.org/datasets/d22a6decd9b147fd9040f793082b219b/explore>
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), pp. 3718-3720.
- GDP by Country (2022). Last access: December 15, 2023, <https://www.worldometers.info/gdp/gdp-by-country/>
- Global Carbon Atlas (2024). Last access: July 31, 2024, <https://globalcarbonatlas.org/emissions/carbon-emissions/>
- Global Climate Risk Index (2021). Last access: December 15, 2023, <https://reliefweb.int/report/world/global-climate-risk-index-2021>.
- Leiserowitz, A., Carman, J., Buttermore, N., Wang, X., Rosenthal, S., Marlon, J., Mulcahy, K. (2021): *International Public Opinion on Climate Change*. New Haven, CT: Yale Program on Climate Change Communication and Facebook Data for Good .
- Leiserowitz, A., Carman, J., Buttermore, N., Neyens, L., Rosenthal, S., Marlon, J., Schneider, J., Mulcahy, K. (2022): *International Public Opinion on Climate Change*, New Haven, CT: Yale Program on Climate Change Communication and Data for Good at Meta.
- Our World in Data (2024). Last access: July 31, 2024, <https://ourworldindata.org/>
- Penn World Table (2024). Last access: July 31, 2024, <https://www.rug.nl/ggdc/productivity/pwt/?lang=en>
- R Core Team (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <http://www.R-project.org>
- Standardized World Income Inequality Database (2024). Last access: July 31, 2024, <https://fsolt.org/swiid/>