

PROGRAMME and BOOK of ABSTRACTS

JOCLAD 2025

3 - 5 APRIL

PORTO, PORTUGAL



Associação Portuguesa de
Classificação e Análise de Dados

XXXII MEETING OF THE PORTUGUESE ASSOCIATION FOR CLASSIFICATION AND DATA ANALYSIS
XXXII JORNADAS DE CLASSIFICAÇÃO E ANÁLISE DE DADOS



Programme and Book of Abstracts

XXXII Meeting of the Portuguese Association for Classification and Data Analysis (JOCLAD 2025)

Editors: M. Rosário Oliveira, Bruno de Sousa, Jorge Cadima, Maria Eduarda Silva, Nuno Moniz, Paula Brito, Susana Faria

Publisher: CLAD

Printed: Statistics Portugal

ISBN 978-989-35097-2-2

Depósito legal: 544678/25

Number of copies: 125

3 April, 16:40 - 17:00, Room 13.2

Clustering health data time series with the generalized affinity coefficient

Ana Paula Nascimento¹, Mónica Vieira¹, Brígida Mónica Faria², Alexandra Oliveira², Cristina Prudêncio¹, Helena Bacelar-Nicolau³

¹ RISE-Health, Center for Translational Health and Medical Biotechnology Research (TBIO), E2S-Polytechnic of Porto, Porto, Portugal, ananascimento@ess.ipp.pt, mav@ess.ipp.pt, cprudencio@ess.ipp.pt

² Artificial Intelligence and Computer Science Laboratory (LIACC member of LASI), University of Porto, Porto, Portugal, E2S-Polytechnic of Porto, Porto, Portugal, monica.faria@ess.ipp.pt, aao@ess.ipp.pt

³ Faculty of Psychology, University of Lisbon (FPUL), Lisboa, Portugal, Institute of Environmental Health, Faculty of Medicine, University of Lisbon (ISAMB-FMUL), Lisboa, Portugal, hbacelar@psicologia.ulisboa.pt

Diabetes *mellitus* (DM) is one of the non-communicable diseases. The behavior of Disability-Adjusted Life Years (DALYs) over time, in the context of DM, for each country was modeled by ARIMA models [3]. Agglomerative hierarchical cluster analysis was employed using the associated distance of the generalized Affinity coefficient to measure the dissimilarity between ARIMA models [2]. The hierarchy shows three country clusters based on DALY trends over the years.

Keywords: diabetes, DALYs, ARIMA models, affinity coefficient, distance

Public Health focuses on studying and preventing diseases, promoting longer lifespans, and improving quality of life through coordinated efforts and informed decision-making. The primary health metric used by the Global Burden of Disease (GBD) study is the disability-adjusted life year (DALY) [1]. This metric serves as a comprehensive tool to evaluate the effects of diseases, injuries, and risk factors. It merges two key components: years of life lost (YLL) due to early mortality and years lived with disability (YLD), presenting a complete picture of the overall health impact of a specific condition. By synthesizing mortality and morbidity data, DALYs offer deeper insights into the actual burden a disease places on a population [3]. When analyzing the health data for the European Region it is clear to see that non-communicable diseases play a big role in the overall burden of disease [3]. Diabetes mellitus (DM) is one of the most severe non-communicable diseases and it can be prevented with a healthy diet and physical exercise, since obesity and overweight are factors that contribute to the increase in diabetes' prevalence. Analyzing data regarding DM is essential in order to define public and health policies. The behavior of DALYs over time, in the context of DM across several countries constitutes time series data. This time series

data can be represented by Autoregressive Integrated Moving Average (ARIMA) models. The ARIMA models that best fit each time series were estimated [3]. Analyzing these time series data is essential for identifying geographic and demographic patterns, offering deeper insights into health disparities across populations [3]. Therefore, agglomerative hierarchical cluster analysis was employed to explore patterns and relationships among various countries in the context of DM using the estimated ARIMA models. To measure the dissimilarity among different countries in the context of DM the associated distance of the generalized Affinity coefficient between ARIMA models is used [2]. The resulting hierarchy reveals three main clusters of countries, as shown in Figure 1.

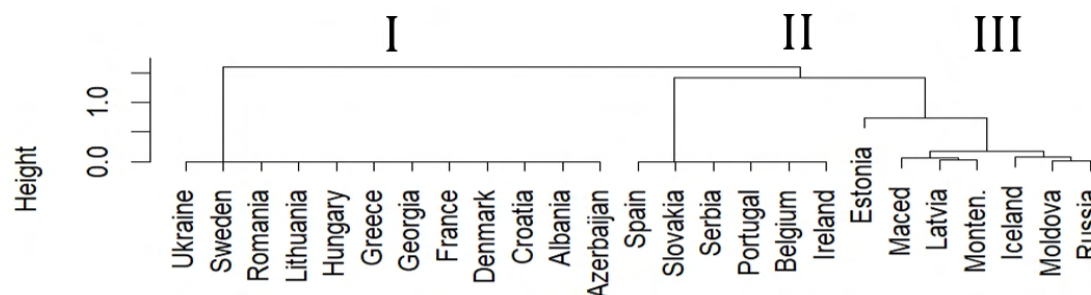


Figure 1: Dendrogram based on the square root of $2(1-\text{Affinity})$ distance and average linkage aggregation criterion between countries in DM context

I - The values in this cluster show an upward trend over the years with fluctuations.
 II - The values in this cluster show an upward trend over the years with a small fluctuation.
 III - The values in this cluster show an upward trend over the years without fluctuations.
 These clusters present the similarities between DALYs rate values from 1990 to 2019 per country and it can be seen that for all groups, there is an increase in DALYs values over the years, indicating a decrease in life-span. However, in Group I, there are periods where these values decrease, suggesting phases of increased life-span, while in Group III, DALYs values consistently increased over the years. These results may help health authorities in the communication between countries and implementation of health policies in order to promote better and longer life-span, as well as economic benefits to different countries.

References

- [1] C. P. Benziger, G. A. Roth, and A. E. Moran. The global burden of disease study and the preventable burden of ncd. *Global Heart*, 11(4):393, Dec 2016.
- [2] A. P. Nascimento, A. Oliveira, B. M. Faria, R. Pimenta, M. Vieira, C. Prudêncio, and H. Bacelar-Nicolau. Affinity coefficient for clustering autoregressive moving average models. *Comp. and Mathematical Methods*, 2024:5540143, 2024.
- [3] C. Vinhal. *Comparing Time Series Forecasting Models for Health Indicators: A Clustering Analysis Approach*. E2S, Polytechnic of Porto, RECIPP, Porto, 2024, to appear.