Instituto Nacional de Estatística
Statistics Portugal

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

The Evolution of Immigrant Groups in Luxembourg

A Symbolic Data Analysis Approach

Catarina Campos Silva[1], Paula Brito[2], Pedro Campos[3]

[1] FEP, University of Porto

[2] FEP, University of Porto & LIAAD INESC TEC, Portugal

[3] FEP, University of Porto & LIAAD INESC TEC & Statistics Portugal, Portugal

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

The conference is partly financed by the European Union

Work carried out through a partnership between the FEP (Faculty of Economics of the University of Porto) and Statec (National Institute of Statistics and Economic Studies of the Grand Duchy of Luxembourg) under the EMOS program.

U.PORTO
FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

STATEC

EUROPEAN MASTER IN OFFICIAL STATISTICS EMOS

The conference is partly financed by the European Union

Contextualization

Immigration in Luxembourg dates back to the late 19th.

More than 50% of people born abroad.

Main reason for immigration: Employment.

Luxembourgish labour market highly internationalised.

Objectives

1. What differentiates the most the different immigrant groups in the labour market?
2. Split the population into homogeneous groups and characterize their profiles in the labour market.
3. Portray the groups' changes over time.

This allows the **identification of disparities** in employment and earnings which may help identifying areas that may require additional support.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

Data

Data Source: Luxembourgish Labour Force Survey

Period covered: 2014 to 2022

Variables: EU-LFS variables + two variables from the National LFS

Cohort: the employees (90% of the labour force)

Methodology

Symbolic Data Analysis

A framework concerned with the analysis of data that exhibit inherent variability.

Cluster Analysis

To group the Symbolic Objects homogeneously.

Heuristic Identification of Noisy Variables algorithm

To select the variables with maximal cluster information.

Monitoring the Evolution of Clusters (MEC) Framework

To portray changes over time.

The conference is partly financed by the European Union

Data Preparation

1. **Pre-processing** micro-data: discretization, handling missing values,...

2. Variables **YEARESID**[1] and **COUNTRYB**[2] were selected in order to aggregate microdata.

3. **21 symbolic objects** were created.

4. Primarily described by **16 modal variables**.

Extract of a symbolic data table for the year 2022

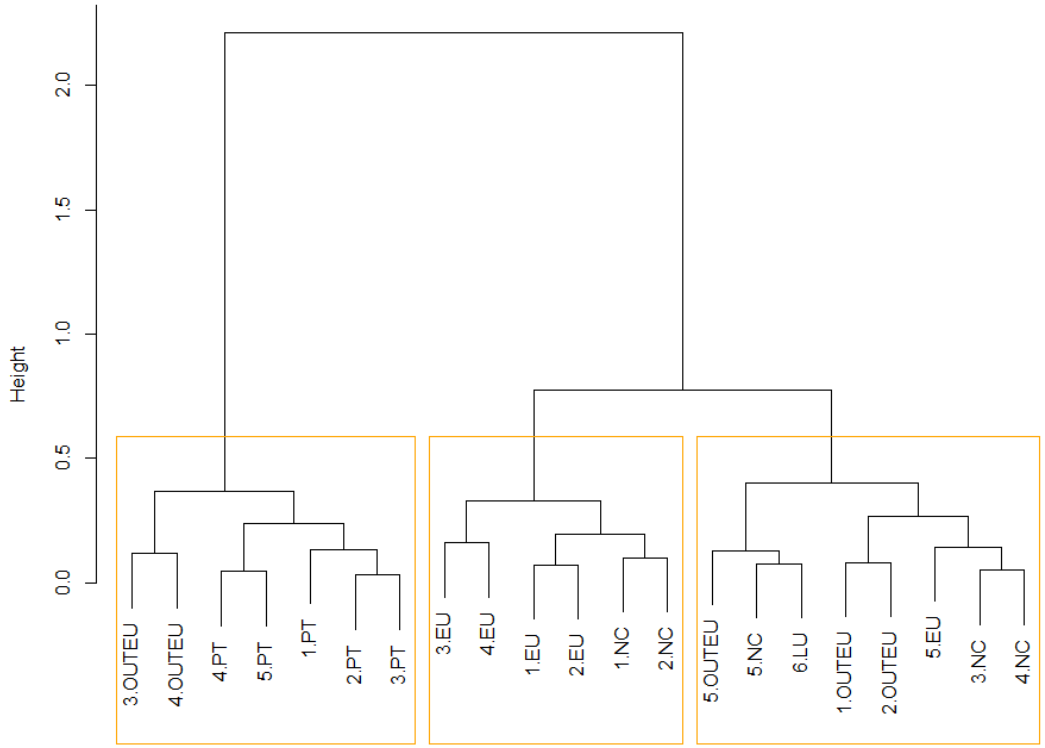
	$H	DEGURBA	1	2	3	$H	TELEARB	1	2
1.EU	$H	3	0.553	0.250	0.197	$H	2	0.671	0.329
1.NC	$H	3	0.360	0.297	0.342	$H	2	0.559	0.441
1.OUTEU	$H	3	0.505	0.351	0.144	$H	2	0.629	0.371
1.PT	$H	3	0.258	0.548	0.194	$H	2	0.194	0.806
2.EU	$H	3	0.582	0.227	0.191	$H	2	0.755	0.245
2.NC	$H	3	0.403	0.371	0.226	$H	2	0.556	0.444
2.OUTEU	$H	3	0.325	0.423	0.252	$H	2	0.447	0.553
2.PT	$H	3	0.188	0.562	0.250	$H	2	0.250	0.750
3.EU	$H	3	0.430	0.289	0.281	$H	2	0.659	0.341
3.NC	$H	3	0.297	0.338	0.366	$H	2	0.538	0.462
3.OUTEU	$H	3	0.194	0.510	0.296	$H	2	0.276	0.724
3.PT	$H	3	0.028	0.697	0.275	$H	2	0.073	0.927
4.EU	$H	3	0.382	0.324	0.294	$H	2	0.647	0.353
4.NC	$H	3	0.333	0.300	0.367	$H	2	0.575	0.425
4.OUTEU	$H	3	0.143	0.520	0.337	$H	2	0.245	0.755
4.PT	$H	3	0.047	0.581	0.372	$H	2	0.047	0.953
5.EU	$H	3	0.340	0.298	0.362	$H	2	0.532	0.468
5.NC	$H	3	0.106	0.440	0.454	$H	2	0.333	0.667
5.OUTEU	$H	3	0.192	0.449	0.359	$H	2	0.231	0.769
5.PT	$H	3	0.066	0.497	0.437	$H	2	0.093	0.907
6.LU	$H	3	0.084	0.398	0.518	$H	2	0.237	0.763

[1]Duration of stay in Luxembourg in completed years. 1: 0 ≤ YEARESID < 4; 2: 4 < YEARESID ≤ 8; 3: 8 < YEARESID ≤ 15; 4: 15 < YEARESID ≤ 24; 5: YEARESID > 24; 6: Born in Luxembourg.
[2]Place of Birth. LU: From Luxembourg; PT: From Portugal; NC: From a neighbour country; EU: From an EU country (excluding above); OUTEU: From a non-EU country.

The conference is partly financed by the European Union

Results

1. **For each year the HCLUST algorithm was applied.**
2. Based on the dendrograms and three validity indices - Silhouette, Index G2 and Index G3 - a partition into *k* clusters was selected.
3. Cluster descriptions were obtained.
4. Cluster Monitoring.

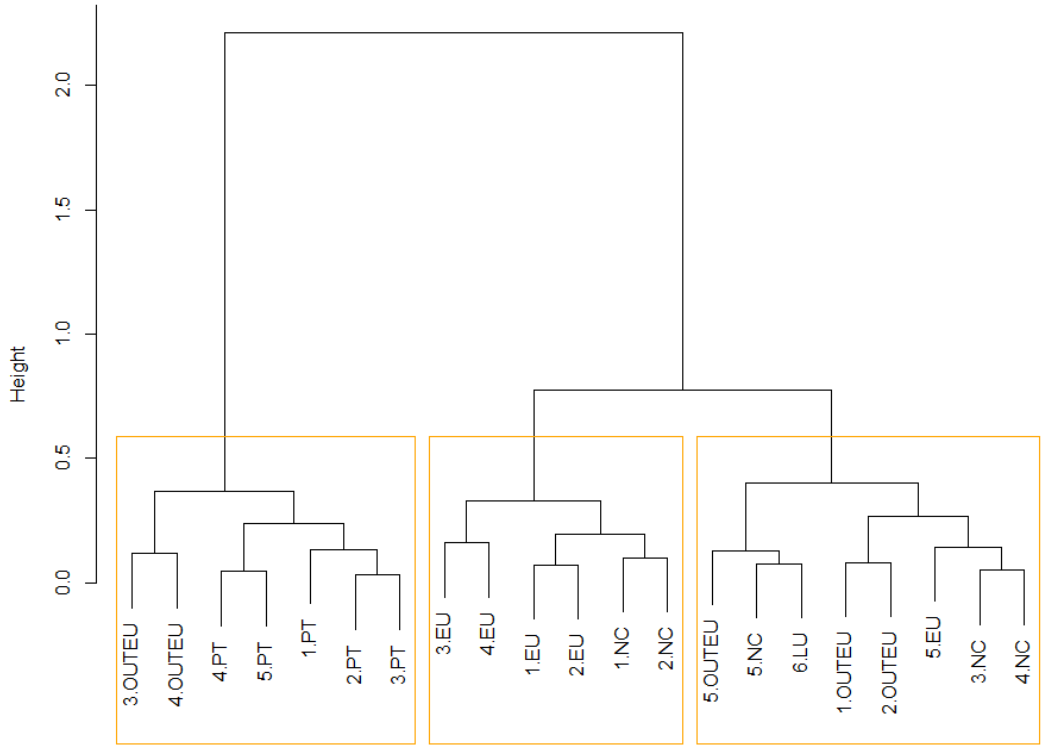


Dendrogram 2015

The conference is partly financed by the European Union

Results

1. For each year the HCLUST algorithm was applied.
2. **Based on the dendrograms and three validity indices - Silhouette, Index G2 and Index G3 - a partition into *k* clusters was selected.**
3. Cluster descriptions were obtained.
4. Cluster Monitoring.



Dendrogram 2015

Results

1. For each year the HCLUST algorithm was applied.
2. Based on the dendrograms and three validity indices - Silhouette, Index G2 and Index G3 - a partition into *k* clusters was selected.
3. **Cluster descriptions were obtained.**
4. Cluster Monitoring.

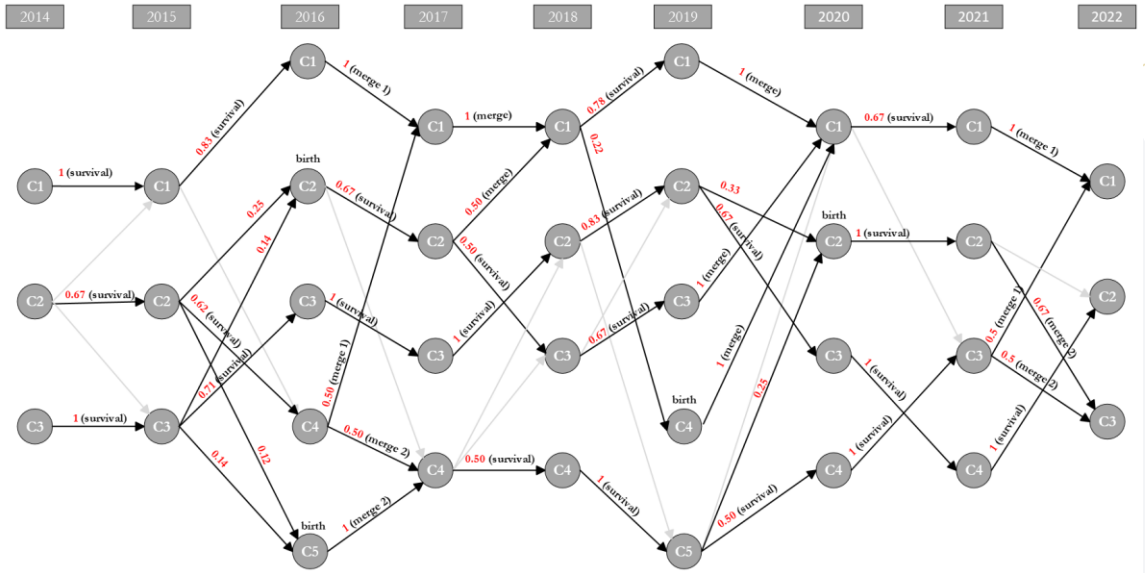
		Cluster		
		1	2	3
TELEARB	Yes	0.597	0.131	0.264
	No	0.403	0.869	0.736
ISCO4D	White-collar high-skilled professions	0.831	0.267	0.573
	White-collar low-skilled professions	0.094	0.169	0.178
	Blue-collar high-skilled professions	0.018	0.161	0.078
	Blue-collar low-skilled professions	0.056	0.403	0.171
HATLEVEL	High School	0.197	0.786	0.541
	Bachelor	0.224	0.096	0.201
	Master/PhD	0.579	0.119	0.258
NACE3D	Agriculture	0.000	0.003	0.003
	Industry	0.030	0.047	0.053
	Construction	0.014	0.204	0.059
	Trade, transport and hospitality	0.164	0.228	0.174

Extract of Cluster Descriptions - 2022

The conference is partly financed by the European Union

Results

1. For each year the HCLUST algorithm was applied.
2. Based on the dendrograms and three validity indices - Silhouette, Index G2 and Index G3 - a partition into *k* clusters was selected.
3. Cluster descriptions were obtained.
4. **Cluster Monitoring.**



Cluster Monitoring Graph 2014-2022

The conference is partly financed by the European Union

Results

- The objects were getting closer until the Brexit and the Pandemic.
- 7/9 years the Portuguese groups are together. Sometimes they are alone, other times they are together with some OUTEU groups.
- 1.EU, 2.EU, 3.EU, 4.EU, 1.NC, 2.NC, 3.NC, 4.NC groups tend to be together in the same cluster
- 5.NC and 6.LU are always in the same cluster. The 5.EU group is also in the same cluster as these two groups in 7 out of 9 years
- OUTEU groups are more spread out across clusters and move more.

The conference is partly financed by the European Union

Final Considerations

- The Luxembourgish population was split and different profiles were identified.
- Some of the conclusions are in line with existing research, while others were new. For example, the importance of the degree of urbanization to split the population.
- Use of a methodology with advantages in Official Statistics that is not yet widely used.
- Accounting for time which revealed some changes that might be occurring.

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

2024 ESTORIL - PORTUGAL

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union