Multiple factor analysis to two-way contingency table to compare residential and geographical trajectories

International conference on Correspondence Analysis and Related Methods: CARME 2011

Elisabeth Morand-Bénédicte Garnier-Catherine Bonvalet

Institut National d’Etudes Démographiques
Service Méthodes Statistiques

Rennes, February 10, 2011
INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Componant Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Componant Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Institut National d’Etudes Démographiques

Population and depopulation of Paris (Catherine Bonvalet, 1986)
Retrospective survey Residential history of a generation
Multiple factor analysis to two-way contingency table to compare residential and geographical trajectories

CARME 2011

INTRODUCTION

Survey description

Aims

MATERIALS AND METHODS

RESULTS

CONCLUSION

Références

Biographic and Retrospective

► 1,987 people aged between 50 and 59 (generation 1926-1935) living in the Paris Region

► Sociodemographic characteristics of the population at the time of the survey (1986)

► Each inhabitant is asked about all the dwellings he has lived for a significant residential period (more than a year) during the period before the survey
Geographical life courses in The Paris Region (Ile de France)

4 locations: Paris (PA) - Inner Suburb (IS) - Outer Suburb (OS) - Outside Ile de France (OIDF)

[Map of Ile de France showing different areas and numbers (75, 92, 93, 94, 91, 77, 95, 78)]

- 75: Paris
- 92, 93, 94: Inner Suburb
- 77, 78, 91, 95: Outer Suburb
Residential life courses

4 tenure statuses

- Tenant (T)
- Living in a Council house or 1948 Housing Law (CL)
- Owner (Ow)
- Other (Ot)
Plan : 1 - INTRODUCTION

INTRODUCTION

Survey description
Aims

MATERIALS AND METHODS

Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS

Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Aims

1,987 residents and 7701 dwellings: these data allow us to reconstitute the residential and the geographical trajectories of Paris Region inhabitants.
Aims

The residential trajectory is defined as the succession of tenure statuses during a time interval until the survey.
So we can:

- compare residential and geographical trajectories;
- observe links between residential and geographical trajectories.
Aims

The residential trajectory is defined as the succession of tenure statuses during a time interval until the survey.
So we can:

- compare residential and geographical trajectories
- observe links between residential and geographical trajectories
Aims

The residential trajectory is defined as the succession of tenure statuses during a time interval until the survey. So we can:

- compare residential and geographical trajectories
- observe links between residential and geographical trajectories

The aim is to group together inhabitants according to their residential and geographical trajectories taking into account each specificity.
Aims

The residential trajectory is defined as the succession of tenure statuses during a time interval until the survey. So we can:

- compare residential and geographical trajectories
- observe links between residential and geographical trajectories

The aim is to group together inhabitants according to their residential and geographical trajectories taking into account each specificity.
Plan: 2 - MATERIALS AND METHODS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Qualitative Harmonic Analysis (QHA)

- The aim is to study event-history data i.e for each individual the chronological succession of statuses he has lived. For example the succession of tenure status: owner, tenant, social housing and other
Qualitative Harmonic Analysis (QHA)

- The aim is to study event-history data i.e for each individual the chronological succession of statuses he has lived. For example the succession of tenure status: owner, tenant, social housing and other
- These data can be seen as \( n \) trajectories of a qualitative process \( X_j \) in time continuous and finished numbers of statuses (\( m \)). The name of the method results from the fact that we are decomposing each trajectory in an harmonical weighted sum
Qualitative Harmonic Analysis (QHA)

- The aim is to study event-history data i.e for each individual the chronological succession of statuses he has lived. For example the succession of tenure status: owner, tenant, social housing and other.
- These data can be seen as \( n \) trajectories of a qualitative process \( X_j \) in time continuous and finished numbers of statuses \( (m) \). The name of the method results from the fact that we are decomposing each trajectory in an harmonical weighted sum.
- We make the hypothesis that each individual is observed on the same period \([S, T]\). This interval is cut in \( p \) sub-period \([t_{k-1}; t_k]\) \((t_0 = T_1 \text{ et } t_p = T_2)\).
Harmonic matrix

A good approximation of Qualitative Harmonic Analysis consists in performing a correspondence analysis on a table constituted by $y_{ijk}$, in which we find the duration proportion of each individual $i$ spent in each state $j$ during each time interval.

A correspondence analysis is performed on this data called Harmonic matrix.

$$[t_0; t_1[, = [S; t_1], \ldots \quad [t_{p-1}; t_p] = [t_{p-1}; T]$$

<table>
<thead>
<tr>
<th>status$_1$</th>
<th>status$_j$</th>
<th>status$_m$</th>
<th>\ldots</th>
<th>status$_1$</th>
<th>status$_j$</th>
<th>status$_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>\ldots</td>
<td>\ldots</td>
<td>1</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>\vdots</td>
<td></td>
<td>$y_{ijk}$</td>
<td></td>
<td>\vdots</td>
<td></td>
<td>\vdots</td>
</tr>
</tbody>
</table>

$y_{ijk}$
Plan: 2 - MATERIALS AND METHODS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Correspondance analysis is a Principal Componant Analysis (PCA) on the table X with general term

\[ x_{ij} = \frac{f_{ij}}{f_i \sqrt{f_j}} \]

and the row weight equal to \( f_i \).

So in our case, because of harmonic matrix construction, all rows have the same weight: \( f_i = p/n \).
2 trajectories

We have 2 trajectories but individual (here the inhabitants) are the same in both data set
2 QHA to perform = 2 PCA to perform
INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Componant Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Multiple Factor Analysis : MFA

- Multiple Factor Analysis (MFA; Escofier & Pagès 1988-1998) deals with table in which a set of individuals is described by 2 or more sets of variables
- The core of the MFA consists in performing a weighted Principal Components Analysis (PCA) on the complete data table $X=(X_1, X_2)$
- Each variable of the group $j$ ($j=1,2$) is weighted by $\lambda_1 j$ (the first eigenvalue of PCA applied to set $j$)

The weighting of the variables as made in MFA allows to balance the role of each variables group in the analysis
Plan: 2 - MATERIALS AND METHODS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Classification

Factor analysis is followed by an Ascending Hierarchical Classification on individuals by selecting the common factors with the highest inertia.
Plan: 3 - RESULTS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
The residents of the Paris Region questioned at the moment of the survey are aged between 50 and 60.

We need each trajectories to have the same length, so we took trajectories of each individual between 25 and 50 years

To describe the event history data we have to divide the period between 25yo and 50yo into a finite number of intervals
A strategic choice is the number of intervals and the length of each of them.

**Sub-period**

Here 10 groups of ageing period are used:
25-26 yo, 27 yo, 28 yo, 29yo, 30-31 yo, 32-33 yo, 34-35 yo, 36-39 yo, 40-44yo, 45-50 yo

numerous intervals at the beginning period to perceive changes in the localisation and tenure during the period where there are changes;

less interval in the final period of the trajectory concerning less changes in the trajectories.
### Geographical trajectories

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Residential trajectories

<table>
<thead>
<tr>
<th>id</th>
<th>Ow.25.26</th>
<th>CL.25.26</th>
<th>Te.25.26</th>
<th>Ot.25.26</th>
<th>Ow.27</th>
<th>CL.27</th>
<th>Te.27</th>
<th>Ot.27</th>
<th>Ow.28.29</th>
<th>CL.28.29</th>
<th>Te.28.29</th>
<th>Ot.28.29</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Plan : 3 - RESULTS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Componant Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
To be out suburb or not

first map

Dim 1 (14.67%)

Dim 2 (12.91%)

-3 -2 -1 0 1 2 3
-2 -1 0 1 2 3

OldF.46-50
OldF.41-45
Ot.25-26
Os.25-26
OIdF.41-45
OIdF.46-50
Ow.25-26
CL.25-26
Ot.25-26
CL.41-45
Te.41-45
Ot.41-45
CL.46-50
Te.46-50
Ot.46-50
PA.46-50
PA.25-26
CL.46-50
CL.41-45
CL.25-26

Dim 1 (14.67%)

Dim 2 (12.91%)

-3 -2 -1 0 1 2 3
-2 -1 0 1 2 3

first map

Dim 1 (14.67%)

Dim 2 (12.91%)

-3 -2 -1 0 1 2 3
-2 -1 0 1 2 3

OldF.46-50
OldF.41-45
Ot.25-26
Os.25-26
OIdF.41-45
OIdF.46-50
Ow.25-26
CL.25-26
Ot.25-26
CL.41-45
Te.41-45
Ot.41-45
CL.46-50
Te.46-50
Ot.46-50
PA.46-50
PA.25-26
CL.46-50
CL.41-45
CL.25-26
To be out suburb or not

Time is such a strong structure that it is underlined by the first axes even if: events are grouped by period and time ordered is not taken into account by the analysis,
To be out suburb or not

first map

PARIS

Out Suburb Owner

Dim 1 (14.67%)

Dim 2 (12.91%)
Inner suburb

second map

Dim 3 (11.13%)
Dim 4 (10.32%)
PA.25-26
IS.25-26
OS.25-26
OIdF.25-26
OIdF.41-45
PA.46-50
IS.46-50
OS.46-50
OIdF.46-50
Ow.25-26
CL.25-26
Te.25-26
Ot.25-26
Ot.41-45
Ow.46-50
CL.46-50
Te.46-50
Ot.46-50
Inner suburb

second map

Dim 3 (11.13%)

Dim 4 (10.32%)

PA.25-26
IS.25-26
OS.25-26
OIdF.25-26
OIdF.41-45
PA.46-50
IS.46-50
OS.46-50
OIdF.46-50

Ow.25-26
CL.25-26
Te.25-26
Ot.25-26
Ot.41-45
Ow.46-50
CL.46-50
Te.46-50
Ot.46-50
Ot
Individual trajectories
Individual trajectories

From 25 yo to 50 yo 0938 stay owner and out suburb
Individual trajectories

Multiple factor analysis to two-way contingency table to compare residential and geographical trajectories

CARME 2011

INTRODUCTION

MATERIALS AND METHODS

RESULTS

Analysis period, time discretisation

Factor analysis

Classification

CONCLUSION

Références
Plan: 3 - RESULTS

INTRODUCTION
Survey description
Aims

MATERIALS AND METHODS
Qualitative Harmonic Analysis (QHA)
Correspondance analysis is a Principal Component Analysis
Multiple Factor Analysis
Classification

RESULTS
Analysis period, time discretisation
Factor analysis
Classification

CONCLUSION
Classification

The Ascending Hierarchical Classification based on the first 6 factors which carry 66% of the inertia
Based on the classification, 5 groups were obtained:

<table>
<thead>
<tr>
<th>Group</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>class 1</td>
<td>21</td>
</tr>
<tr>
<td>class 2</td>
<td>15</td>
</tr>
<tr>
<td>class 3</td>
<td>25</td>
</tr>
<tr>
<td>class 4</td>
<td>20</td>
</tr>
<tr>
<td>class 5</td>
<td>19</td>
</tr>
</tbody>
</table>
Classification
Classification
Classification
Discussion

- Exploratory aspect
- Considering partial and missing answers
- Exploration of complexes trajectories
Perspectives

▶ To continue the exploration: Link to family story (union formation, birth, departure of children)
▶ Complementarity with other methods
▶ Fields of application: such as conjugal and parental trajectories, transition to adulthood
References I


