Multidimensional Scaling in R: SMACOF

Patrick Mair
Department of Statistics and Mathematics
WU Vienna University of Economics and Business

Jan de Leeuw
Department of Statistics
University of California, Los Angeles (UCLA)

Eva Lienbacher
Institute for Retailing and Marketing
WU Vienna University of Economics and Business
Content

• Basics of MDS and SMACOF
• SMACOF implementation in R
• Symmetric SMACOF
• Spherical SMACOF
• Rectangular SMACOF
• 3-Way SMACOF
Multidimensional Scaling (MDS)

MDS: Family of data-analytic methods which represent distances between objects in a low-dimensional space.

• Torgerson (1952): Classical scaling approach introduced to Psychometrics.
• Shepard (1962): Non-metric MDS.
• Kruskal (1964): Stress, reduction of dimensions.
MDS Workflow

We have the following steps of analysis:

1. Input structure: Dissimilarity (distance) matrix.

2. Computation: Optimize target function (stress).

3. Output: Configurations in low-dimensional space.

4. Visualization: Configuration plot, goodness-of-fit plots.
Distance Matrix and Computation

R> islanddist

<table>
<thead>
<tr>
<th></th>
<th>Crete</th>
<th>Euboea</th>
<th>Lesbos</th>
<th>Rhodes</th>
<th>Kefalonia</th>
<th>Chios</th>
<th>Corfu</th>
<th>Lemnos</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crete</td>
<td>-</td>
<td>379.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euboea</td>
<td>379.6</td>
<td></td>
<td>455.90</td>
<td>215.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesbos</td>
<td>455.9</td>
<td>215.70</td>
<td></td>
<td>292.00</td>
<td>438.90</td>
<td>364.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodes</td>
<td>379.6</td>
<td>455.9</td>
<td>292.00</td>
<td></td>
<td>521.90</td>
<td>307.7</td>
<td>518.7</td>
<td>702.90</td>
<td></td>
</tr>
<tr>
<td>Kefalonia</td>
<td>379.6</td>
<td>455.9</td>
<td>292.00</td>
<td>521.90</td>
<td></td>
<td>175.30</td>
<td>560.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chios</td>
<td>367.3</td>
<td>177.00</td>
<td>88.62</td>
<td>304.00</td>
<td>483.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corfu</td>
<td>679.2</td>
<td>389.80</td>
<td>569.50</td>
<td>822.40</td>
<td>175.30</td>
<td>560.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemnos</td>
<td>524.0</td>
<td>190.80</td>
<td>124.90</td>
<td>481.10</td>
<td>453.10</td>
<td>180.6</td>
<td>471.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samos</td>
<td>325.7</td>
<td>259.20</td>
<td>162.40</td>
<td>205.70</td>
<td>555.70</td>
<td>99.65</td>
<td>648.9</td>
<td>275.60</td>
<td></td>
</tr>
<tr>
<td>Naxos</td>
<td>210.0</td>
<td>206.70</td>
<td>246.30</td>
<td>246.30</td>
<td>457.30</td>
<td>157.5</td>
<td>578.3</td>
<td>318.50</td>
<td></td>
</tr>
<tr>
<td>Zakynthos</td>
<td>477.6</td>
<td>295.30</td>
<td>510.10</td>
<td>670.20</td>
<td>49.53</td>
<td>467.4</td>
<td>224.7</td>
<td>455.70</td>
<td></td>
</tr>
<tr>
<td>Thasos</td>
<td>607.5</td>
<td>247.30</td>
<td>219.20</td>
<td>578.00</td>
<td>451.20</td>
<td>276.8</td>
<td>433.2</td>
<td>96.93</td>
<td></td>
</tr>
</tbody>
</table>

R> res.island <- smacofSym(islanddist)
SMACOF in R
R Implementation: SMACOF package

The R Project for Statistical Computing

- R is an open source software environment for statistical computing and graphics
- http://www.R-project.org
- 1938 packages available

The smacof package

- CRAN: http://CRAN.R-project.org
- PsychoR: http://r-forge.r-project.org/projects/psychor.
Symmetric SMACOF

- Distance matrix $\Delta$ of dimension $n \times n$ with elements $\delta_{ij}$.
- Problem to solve: Locate points (configurations) in a $p$-dimensional space such that the distances $d_{ij}(X)$ between the points approximate $\delta_{ij}$.
- Configuration distances:

$$d_{ij}(X) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}$$

- Minimize stress (Majorization; de Leeuw, 1977):

$$\sigma(X) = \sum_{i<j} w_{ij} (\delta_{ij} - d_{ij}(X))^2 \rightarrow \min!$$
Example 1: Signs of the Zodiac
Example 1: Signs of the Zodiac

(Thanks to Paul Eigenthaler from the Institute of Astronomy, University of Vienna, for providing the distances.)

R> resall <- smacofSym(stardist, ndim = 2)
R> resall
Call: smacofSym(delta = stardist, ndim = 2)

Model: Symmetric SMACOF
Number of objects: 120

Metric stress: 7.817851e-05
Number of iterations: 818
Spherical SMACOF

Restrictions on the configurations (*weakly constrained MDS*).

\[ x_i' \Lambda x_i + 2x_i' \beta + \gamma = 0 \]

- \( \mathbb{R}^2 \): circle, ellipse, hyperbola, parabola.
- \( \mathbb{R}^3 \): sphere, ellipsoid, hyperboloid, paraboloid, cylinder.
- Optimization: Primal and dual methods available.
Example 2: Trading Volume

R> tradedist

<table>
<thead>
<tr>
<th></th>
<th>North America</th>
<th>South America</th>
<th>Europe</th>
<th>Commonwealth</th>
<th>Africa</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>South America</td>
<td>367.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>362.1</td>
<td>963.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commonwealth</td>
<td>1113.3</td>
<td>1136.6</td>
<td>672.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>1030.1</td>
<td>1121.0</td>
<td>834.1</td>
<td>1141.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle East</td>
<td>1015.3</td>
<td>1135.8</td>
<td>888.1</td>
<td>1128.3</td>
<td>1111.3</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>40.8</td>
<td>976.8</td>
<td>1.0</td>
<td>1009.9</td>
<td>977.0</td>
<td>...</td>
</tr>
</tbody>
</table>

R> res.trade <- smacofSphere.dual(tradedist, ndim = 3, itmax = 2000)
R> plot3d(res.trade, sphere = TRUE)
SMACOF in R

Configuration Plot

Commonwealth

Middle East

Europe

Asia

North America

South America

Africa

Dimension 1

Dimension 2

Dimension 3
Rectangular SMACOF (Unfolding)

Rectangular $n_1 \times n_2$ preference matrix $\Delta$.

- Stress becomes

$$
\sigma(X) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij} (\delta_{ij} - d_{ij}(X_1, X_2))^2 \rightarrow \min!
$$

- Judge $n_1 \times p$ configuration matrix
- Object $n_2 \times p$ configuration matrix
Example 3: Company Rating

R> head(csr)
   Environment Waste Prevention Organic Products Charity Employee
1      1         2          4          3         5
2      2         1          5          4         3
3      1         5          3          4         2
4      3         1          5          4         2
5      2         3          4          5         1
6      2         1          4          5         3

R> res.csr <- smacofRect(csr)

R> plot(res.csr, xlim = c(-3, 3), joint = TRUE, asp = 1)
SMACOF in R

Joint Configuration Plot

Configurations D1
Configurations D2
Environment
Waste Prevention
Employee
Charity
Organic Products
3-Way SMACOF

SMACOF for individual differences:

- $k = 1, \ldots, K$ separate symmetric distance matrices.
- Data cube, or, in R: List.
Example 4: Wine Tasting

- Ziniel Chardonnay
- Markowitsch Chardonnay
- Krems Chardonnay
- Castel Nova Chardonnay
- Ritinitis Noble Retsina
- Retsina

Criteria: color, smell, taste, fun, overall impression

R> reswine <- smacofIndDiff(winedat, metric = FALSE)

R> plot(reswine, xlim = c(-1, 1))
### Wine Tasting: Descriptives

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Alcohol</th>
<th>Mean</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jurtschitsch Chardonnay</td>
<td>14.99</td>
<td>13.00</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>Ziniel Chardonnay</td>
<td>7.00</td>
<td>12.00</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>Markowitsch Chardonnay</td>
<td>9.99</td>
<td>12.50</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>Ritinitis Noble Retsina</td>
<td>9.99</td>
<td>12.00</td>
<td>4.30</td>
<td></td>
</tr>
<tr>
<td>Retsina</td>
<td>2.99</td>
<td>11.50</td>
<td>4.60</td>
<td></td>
</tr>
<tr>
<td>Krems Chardonnay</td>
<td>5.99</td>
<td>12.50</td>
<td>2.70</td>
<td></td>
</tr>
<tr>
<td>Castel Nova Chardonnay</td>
<td>1.99</td>
<td>12.00</td>
<td>2.80</td>
<td></td>
</tr>
</tbody>
</table>
Additional Models and Options

Each SMACOF variant is implemented in a metric and non-metric way.

- If observed data are ordinal → distances will be ordinal as well → non-metric MDS.
- Various distance measures (e.g. Euclidean, Jaccard, Minkowski, etc.), proxy package in R.
- Estimation: Additional isotone regression step (PAVA).
Additional models and options

Decomposition of the configurations (de Leeuw & Heiser, 1980):

- Linear decomposition $X = ZC$.
- SMACOF function `smacofConstraint()`.

More 3-way options:

- IDIOSCAL (Carrol & Wish, 1974)
- Various other decompositions of the weight matrix.

Goodness-of-fit examination: Shepard diagrams, Stress plots, Residual plots.
References


SMACOF in R

Links and Contact

PsychoR project:

- Website: http://r-forge.r-project.org/projects/psychor
- Next PsychoR topics: isotone optimization, exponential geometric models, homals with splines.

Patrick Mair
Department of Statistics and Mathematics
WU Vienna University of Economics
Augasse 2-6
1090 Vienna

Email: patrick.mair@wu.ac.at
Website: http://statmath.wu.ac.at/~mair

5th Panhellenic Data Analysis Conference 2009-09-11