Smacof at 50: A Manual
Part 8: Homogeneity Analysis with Smacof

Jan de Leeuw - University of California Los Angeles

Started April 13 2024, Version of June 17, 2024

Abstract

smacofVO

Contents

Note: This is a working manuscript which will be expanded/updated frequently. All suggestions for improvement are welcome. All Rmd, tex, html, pdf, R, and C files are in the public domain. Attribution will be appreciated, but is not required. The files can be found at https://github.com/deleeuw in the repositories smacofCode, smacofManual, and smacofExamples.

1 Simultaneous Non-Metric Unfolding

- There are m *variables*.
- Variable j has $k_j > 1$ *categories*.
- There are n *objects*.
- Each object defines a partial order over the categories of each variable.

The problem we analyze in this chapter is to minimize the stress loss function

$$\sigma(X, Y_1, \cdots, Y_m) := \sum_{j=1}^m \sum_{i=1}^n \min_{\hat{d}_i^j \in \Delta_i^j} \sum_{l=1}^{k_j} w_{il}^j(\hat{d}_{il}^j - d(x_i, y_l^j))^2 \tag{1}$$

Note that for each object and variable there are different sets of transformations Δ_j and for each variable there different matrices of *category scores* Y_j, but there is only a single matrix of *object scores* X. Also note that index j, for variables, is sometimes used as a subscript and sometimes as a superscript, depending on what looks best.

If the Δ_i^j contain the zero vector, then the unconstrained minimum of (1) iszero. Collapsing all x_i and all y_l^j into a single point makes all distances zero, and thus makes stress zero. In fact, it is easy to see that a minimum of zero is also possible in the situation where the Δ_i^j contain the set of all constant vectors (or all non-negative constant vectors). Collapse all x_i into a single point, and place all y_l^j for variable j on a sphere around this point. There can be different spheres for different variables. This makes all $d(x_i, y_l^j)$ equal and thus makes stress zero. Some comstraints on X and/or the Y_j is needed to prevent this trivial solution.

2 Homogeneity Analysis

The Gifi System (Gifi (1990), Michailidis and De Leeuw (1998), De Leeuw and Mair (2009)) implements non-linear or non-metric versions of the classical linear multivariate analysis techniques (regression, analysis of variance, canonical analysis, discriminant analysis, principal component analysis) as special cases of *Homogeneity Analysis*, better known as *Multiple Correspondence Analysis*.

In this section we present homogeneity analysis as a technique for minimizing the loss function (1) when the data are $n \times k_j$ *indicator matrices* G_j, with $j = 1, \cdots, M$. This is a non-standard presentation, because usually homogeneity analysis is related to principal component analysis, and not to multidimensional scaling (see, for example, De Leeuw (2014) or De Leeuw (1923)). Indicator matrices are binary matrices, with rows that add up to one or to zero. Thus each row has either a single elements equal to one and the rest zeroes, or all elements equal to zero. Indicator matrices are used to code categorical variables. Rows corresponds with objects (or individuals), columns with the categories (or levels) of a variable. Element g_{il}^j is one if object i is in category l of variable j, and all other elements in row i are zero. If an object is *missing* on variable j then the whole row is zero.

Homogeneity analysis makes joint maps in p dimensions of objects and categories, both represented as points. A joint map for variable j has n object points x_i and k_j category points y_{il}^j. In a homogeneous solution the object points are close to the points of the categories that the objects score in, i.e, to those y_{il}^j for which $g_{il}^j = 1$. If there is only one variable then it is trivial to make a perfectly homogeneous map. We just make sure the object points coincide with their category points. But there are $j > 1$ indicator matrices, corresponding with m categorical variables, and there is only a single set of object scores. The solution is a compromise trying to achieve as much homogeneity as possible for all variables simultaneously.

In loss function (1) applied to homogeneity analysis the sets Δ_i^j are defined in such a way that $\widehat{d}_{il}^j$ is zero if i is in category l of variable j. There are no constraints on the other $\widehat{d}$'s in row i of variable j. Thus for zero loss we want an object to coincide with all m categories it is in. With this definition of the Δ_i^j we have

$$\min_{\widehat{d}_i^j \in \Delta_i^j} \sum_{l=1}^{k_j} w_{il}^j (\widehat{d}_{il}^j - d(x_i, y_l^j))^2 = w_{ij} d_{ij}^2(X, Y), \tag{2}$$

where

$$d_{ij}(X, Y) = \sum_{l=1}^{k_j} g_{il}^j d(x_i, y_l^j), \tag{3}$$

$$w_{ij} = \sum_{l=1}^{k_j} g_{il}^j w_{il}^j. \tag{4}$$

Using indicator matrices we can write loss function (2) as

$$\sigma(X, Y_1, \cdots, Y_m) = \sum_{j=1}^{m} \text{tr } (X - G_j Y_j)' W_j (X - G_j Y_j), \tag{5}$$

The W_j are diagonal matrices with the w_{ij} from (4) on the diagonal.

In homogeneity analysis we minimize (5) using the explicit normalization $X'W_\star X = I$, where $W_\star$ is the sum of the W_j. The solution is given by the singular value equations

$$X\Lambda = W_\star^{-1}\sum_{j=1}^m W_jG_jY_j, \tag{6}$$

$$Y_j = (G_j'W_jG_j)^{-1}G_j'W_jX, \tag{7}$$

where Λ is a symmetric matrix of Lagrange multipliers. Remember that in addition we require $X'W_\star X = I$.

In homals (Gifi (1980), De Leeuw and Mair (2009)) alternating least squares is used to solve the equations (6) and (7). We start with some initial X, then compute the corresponding Y_j using (7), then for these new Y_j we compute a new corresponding X from (6), and so on. Computations are efficient, because only diagonal matrices need to be inverted and matrix multiplication with an indicator matrix is not really multiplication but simply selection of a particular row or column. Alternating least squares becomes reciprocal averaging.

Alternative methods of computation (and interpretation) are possible if we substitute (7) in (6) to eliminate the Y_j and obtain an equation in X only. This gives

$$W_\star X\Lambda = \sum_{j=1}^m W_jG_j(G_j'W_jG_j)^{-1}G_j'W_jX, \tag{8}$$

which is a generalized eigenvalue equation for X. If we substitute (6) in (7) we obtain generalized eigenvalue equations for Y.

$$(G_j'W_jG_j)Y_j\Lambda = \sum_{h=1}^m G_j'W_jW_\star^{-1}W_hG_hY_h. \tag{9}$$

If $k_\star$, the sum of the k_j, is not too large then finding the p largest non-trivial eigenvalues with corresponding eigenvectors from (9) may be computationally efficient. The largest "trivial" eigenvalue is always equal to one, no matter what the G_j and W_j are, and we can safely ignore it. The trivial solutions mentioned in section 1 correspond with this largest eigenvalue.

Homogeneity analysis can be most naturally introduced using the concept of a *star plot*. For variable j we plot k_j category points and n object points in a single joint plot. We then draw a line from each category point to the object points of the objects in that category. This creates k_j groups of lines and points in $\mathbb{R}^p$, and each of these groups is called a *star*. The sum of squares of the line lengths of a star is the loss of homogeneity for category l of variable j, and the total sum of squares of all line lengths in the k_j stars is the loss (5) for variable j. Homogeneity analysis chooses X and the Y_j such that X is normalized by $X'W_\star X = I$ and the stars are as small or as compact as possible, measured by the squared line lengths. For given X the stars are as small as possible by choosing the category points Y_j as the centroids of the object points in the category. That explains the use of the word star, because now the stars really look like stars. Thus, given the choice of the Y_j as centroids, we can also say that homogeneity analysis quantifies the n objcets in such a way that the resulting stars are as small as possible.

3 Loss Function

The smacofHO technique solves the closely related problem in which we do not require, as in homogeneity analysis, that

$$\sum_{l=1}^{k_j} g_{il}^j \hat{d}_{il}^j = 0 \tag{10a}$$

for all i and j, but we impose the weaker condition that for all i and j

$$\sum_{l=1}^{k_j} g_{il}^j \hat{d}_{il}^j \leq \hat{d}_{i\nu}^j \tag{10b}$$

for all $\nu = 1, \cdots, k_j$. In homogeneity analysis the geometric interpretation of loss is that we want objects to coincide with all categories they score in. The geometric interpretation of loss function … is that we want objects to be closer to the categories they score in than to the categories they do not score in.

This can be formalized using the notion of *Voronoi regions*. The Voronoi region of category l of variable j is the polyhedral convex set of all points of $\mathbb{R}^p$ closer to category l than to any other category of variable j. The plot of the the k_j categories of variable j defines k_j Voronoi regions that partition $\mathbb{R}^p$.

Loss function … with Δ defined by … vanishes if for each variable all x_i are in the Voronoi regions of the categories they score in. This condition implies, by the way, that the interiors of the k_j convex hulls of the x_i in a given category are disjoint, and the point clouds can consequently be weakly separated by hyperplanes. Since the category points themselves are in their own Voronoi region the convex hulls of the stars are also disjoint.

3.1 The Guttman Transform

The general majorization theory for MDS with restrictions calls for updates in two steps. In the first step we compute the Guttman transform of the current configuration, and in the second step we project the Guttman transform on the set of constrained configurations.

Configuration updates are alternated with updates of the $\hat{D}_j$. Minimizing loss … over the $\hat{d}_i^j$ is a monotone regression problem for a simple tree order. This is easily solved by using Kruskal's primary approach to ties (Kruskal (1964a), Kruskal (1964b), De Leeuw (1977)).

3.2 Six analyses

There are three options for updating the Y_j.

1. No further constraints on the Y_j.
2. The Y_j are centroids, i.e. $Y_j = (G_j'W_jG_j)^{-1}G_j'W_jX$.
3. The Y_j have rank one, i.e. $Y_j = q_ja_j'$.

There are two options for updating X. Note that X is always constrained to be the same for all variables, i.e. $X_j = X$.

1. No further constraints on X.
2. X is normalized by $X'W_\star X = I$

Note that the centroid constraint on the Y_j and the normalization constraint on X are inspired by homogeneity analysis. The rank-one constraint on Y_j is taken from the Gifi system, where it serves to make homogeneity analysis into a form of non-linear principal component analysis.

3.3 The Unconstrained Case

Suppose there are no constraints on the Y_j and on X (except for the constraint $X_j = X$). The smacof iterations, or Guttman transforms, more or less ignore the fact that we are dealing with a rectangular matrix and use the weights to transform the problem into a symmetric one (as in Heiser and De Leeuw (1979)).

The loss function is

$$\sigma(Z_1, \cdots, Z_m) = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\sum_{k=1}^{n_j} w_{ik}^j(\hat{d}_{ik}^j - d_{ik}(Z_j))^2,$$

with $n_j := n + k_j$ and with Z_j the $n_j \times p$ matrix that stacks X on top of Y_j. The w_{ik}^j are zero for the diagonal $n \times n$ and the diagonal $k_j \times k_j$ block.

To compute the Guttman transform of Z_j we have to solve the partitioned system

$$\begin{bmatrix} R_W & -W \\ -W' & C_W \end{bmatrix}\begin{bmatrix} \tilde{X} \\ \tilde{Y} \end{bmatrix} = \begin{bmatrix} R_B & -B \\ -B' & C_B \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}$$

Since we have to solve this system for each variable separately we forget about the index j here. In … R_W and C_W are the diagonal matrices with row and column sums of $-W$, while R_B and C_B are diagonal matrices with the row and columns sums of the $n \times k_j$ matrix B, which has elements

$$b_{il} = w_{il}\frac{\hat{d}_{il}}{d(x_i, y_l)}$$

Matrices X and Y are the two parts of the current Z that we are updating, while we solve for $\tilde{X}$ and $\tilde{Y}$, the two parts of the Guttman transform. Define

$$\begin{bmatrix} P \\ Q \end{bmatrix} := \begin{bmatrix} R_B & -B \\ -B' & C_B \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}$$

Now $R_W\tilde{X} - W\tilde{Y} = P$ or $\tilde{X} = R_W^{-1}(P + W\tilde{Y})$. Substitute this in $C_W\tilde{Y} - W'\tilde{X} = Q$ to get $C_W\tilde{Y} - W'R_W^{-1}(P + W\tilde{Y}) = Q$ or

$$(C_W - W'R_W^{-1}W)\tilde{Y} = Q + W'R_W^{-1}P$$

We solve … for $\tilde{Y}$and then use $\tilde{X} = R_W^{-1}(P+W\tilde{Y})$. Note that $C_W-W'R_W^{-1}W$is doubly-centered and $Q + W'R_W^{-1}P$is column-centered. As in homogeneity analysis we hope that $k_\star$ is not to big, and we avoid generalized inverses of very large and very sparse matrices.

After computing the Guttman transforms $\tilde{X}_j$ and $\tilde{Y}_j$
we have to project them on the constrained configurations, in this case on the set $X_j = X$. More explicitly we must minimize

$$\sum_{j=1}^m \text{tr } (X-\tilde{X}_j)'R_j(X-\tilde{X}_j) - 2\sum_{j=1}^m \text{tr } (X-\tilde{X}_j)'W_j(Y_j-\tilde{Y}_j)+ \\ \sum_{j=1}^m \text{tr } (Y_j-\tilde{Y}_j)'C_j(Y_j-\tilde{Y}_j) \quad (11)$$

where R_j and C_j are now the diagonal matrices of row and column sums of the $-W_j$.

The stationary equations are

$$Y_j = \tilde{Y}_j - C_j^{-1}W_j'(X-\tilde{X}_j), \quad (12)$$

$$X = \{R_\star\}^{-1}\sum_{j=1}^m \left\{R_j\tilde{X}_j - W_j(Y_j-\tilde{Y}_j)\right\}. \quad (13)$$

We could solve these equations iteratively using alternating least squares. This means using (12) to compute a new $Y$for given X and (13) to compute a new X for given Y. This introduces an infinite inner iteration loop within the main outer iteration loop. We have tried this, and it does not seem to be the way to go. The argument is the same as in our section … on homogeneity analysis.

Alternatively, we can substitute (13) into (12). This gives linear equations in the Y_j, which we can solve. Then use (13) to compute the corresponding optimal X. No inner iterations are necessary. In the same way we can substitute (12) into (13), solve for X and compute the corresponding Y_j. Since in most applications the number of objects n is much larger than the total number of categories $\sum k_j$ the first substitution of (13) into (12) seems the most promising. Again, the argument is the same as in our section … on homogeneity analysis.

$$C_j(Y_j-\tilde{Y}_j) - W_j'R_\star^{-1}\sum_{h=1}^m W_h(Y_h-\tilde{Y}_h) = W_j'(\tilde{X}_j - R_\star^{-1}\sum_{h=1}^m R_h\tilde{X}_h)$$

$$(Y_j-\tilde{Y}_j)-\{V_{22}^j\}^{-1}V_{21}^j\{V_{11}^\star\}^{-1}\sum_{h=1}^m V_{12}^h(Y_h-\tilde{Y}_h) = \{V_{22}^j\}^{-1}V_{21}^j\tilde{X}_j-\{V_{22}^j\}^{-1}V_{21}^j\{V_{11}^\star\}^{-1}\sum_{h=1}^m V_{11}^h\tilde{X}_h$$

$$C_j(Y_j-\tilde{Y}_j) - W_j'R_\star^{-1}\sum_{h=1}^m W_h(Y_h-\tilde{Y}_h) = -W_j\tilde{X}_j + W_j'R_\star^{-1}\sum_{h=1}^m R_h\tilde{X}_h$$

3.4 Normalization of X

Constraint

$$X'R_\star X = I$$

Stationary equation (13) becomes

$$R_\star X\Lambda = \sum_{j=1}^{m} \left\{R_j\tilde{X}_j - W_j(Y_j - \tilde{Y}_j)\right\}$$

If we define

$$P := \sum_{j=1}^{m} \left\{R_j\tilde{X}_j - W_j(Y_j - \tilde{Y}_j)\right\}$$

then, using the symmetric square root, $\Lambda = (P'R_\star^{-1}P)^{\frac{1}{2}}$ and thus $X = R_\star^{-1}P(P'R_\star^{-1}P)^{-\frac{1}{2}}$.

3.5 Centroid Constraints on Y

If we require that $Y_j = (G_j'W_jG_j)^{-1}G_j'W_jX$ then this effectively eliminates the Y_j as variables from the optimization problem and we only have to optimize over X.

$$X - G_jY_j = (I - G_j(G_j'W_jG_j)^{-1}G_j'W_j)X = Q_jX\, d_i^j(X) = \sqrt{\text{tr}\, X'q_i^j\{q_i^j\}'X}$$

3.6 Rank Constraints on Y

As in homals it is possible in smacofHO to impose rank-one restrictions on some or all of the Y_j. This means we require

$$y_l^j = \alpha_l^j y_l$$

Geometrically having all y_l^j on a line through the origin implies that all voronoi boundaries are hyperplanes perpendicular to that line, and consequently all voronoi regions are bounded by two parallel hyperplanes.

$$\begin{aligned}\sum_{j=1}^{m} \text{tr}\,(X-\tilde{X}_j)'R_j(X-\tilde{X}_j) - 2\sum_{j=1}^{m} \text{tr}\,(X-\tilde{X}_j)'W_j(y_ja_j' - \tilde{Y}_j) +& \\ \sum_{j=1}^{m} \text{tr}\,(y_ja_j' - \tilde{Y}_j)'C_j(y_ja_j' - \tilde{Y}_j)&\end{aligned} \tag{14}$$

$$-2a_j'\{(X-\tilde{X}_j)'W_j + \tilde{Y}_j'C_j\}y_j + a_j'a_jy_j'C_jy_j$$

4 Convergence and Degeneracy

5 Utilities

5.1 Object Plot Function

5.2 Category Plots Function

5.3 Joint Plot Function

6 Examples

Presenting examples runs into some difficulties.

6.1 Cetacea

6.2 Senate

6.3 GALO

References

De Leeuw, J. 1923. "Deconstructing Multiple Correspondence Analysis. Essays in Honour of Shizuhiko Nishisato." In *Analysis of Categorical Data from Historical Perspectives*, edited by Eric J. Beh, Rosaria Lombardo, and Jose G. Clavel, 383–407. Springer.

———. 1977. "Correctness of Kruskal's Algorithms for Monotone Regression with Ties." *Psychometrika* 42: 141–44.

———. 2014. "History of Nonlinear Principal Component Analysis." In *The Visualization and Verbalization of Data*, edited by J. Blasius and M. Greenacre. Chapman; Hall.

De Leeuw, J., and P. Mair. 2009. "Homogeneity Analysis in R: the Package homals." *Journal of Statistical Software* 31 (4): 1–21. https://www.jstatsoft.org/v31/i04/.

Gifi, A. 1980. *Niet-Lineaire Multivariate Analyse [Nonlinear Multivariate Analysis]*. Leiden, The Netherlands: Department of Data Theory FSW/RUL.

———. 1990. *Nonlinear Multivariate Analysis*. New York, N.Y.: Wiley.

Heiser, W. J., and J. De Leeuw. 1979. "Metric Multidimensional Unfolding." *Methoden En Data Nieuwsbrief SWS/VVS* 4: 26–50.

Kruskal, J. B. 1964a. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." *Psychometrika* 29: 1–27.

———. 1964b. "Nonmetric Multidimensional Scaling: a Numerical Method." *Psychometrika* 29: 115–29.

Michailidis, G., and J. De Leeuw. 1998. "The Gifi System for Descriptive Multivariate Analysis." *Statistical Science* 13: 307–36.