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 10, 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 all variables.

In this chapter we 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 column scores Y_j, but there is only a single matrix of row 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 zero, then the unconstrained minimum of (1) is clearly zero. Collapsing all x_i and all y_l^j into a single point makes all distances zero, and thus makes stress zero. Some sort of normalization of either X and/or the Y_j is needed to prevent this trivial solution.

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 on a sphere around this point. Or collapse all y_l^j and put the x_i on a sphere. This makes all $d(x_i, y_l^j)$ equal and thus makes stress zero.

1.1 Homogeneity Analysis

The Gifi System (Gifi (1990), Michailidis and De Leeuw (1998), De Leeuw and Mair (2009)) presents 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, also known as Multiple Correspondence Analysis.

In this section we present homogeneity analysis as a special case of minimizing the loss function (1).

The data are a number of indicator matrices $G_1, \cdots, G_m$. Indicator matrices are binary matrices, with rows that add up to one. They are used to code categorical variables. Rows corresponds with objects (or objects), columns with the categories (or levels) of a variable. An element g_{il} is one in row if object i is in category l, and all other elements in row i are zero.

Homogeneity analysis makes a joint maps in p dimensions of objects and categories (both represented as points) in such a way that category points are close to the points for the objects in the category. And, vice versa, objects are close to the category points that they score in.

If there is only one variable then it is trivial to make such a 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 the solution is a compromise trying to achieve homogeneity as well 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 $\hat{d}_{il}^j$ is zero if i is in category l of variable j. There are no constraints on the other $\hat{d}$'s in row i of variable j. Thus for zero loss we want an object to coincide with all m categories it is in. Under this definition of the Δ_i^j we have

$$\min_{\hat{d}_i^j \in \Delta_i^j} \sum_{l=1}^{k_j} w_{il}^j (\hat{d}_{il}^j - d(x_i, y_l^j))^2 = w_{il(i,j)}^j d(x_i, y_{l(i,j)}^j)^2 \tag{2}$$

where the $l(i,j)$ on the right is the index of the category of variable j that object i is in.

Using indicators we can write loss function (2) as

$$\sigma(X, Y_1, \cdots, Y_m) = \sum_{j=1}^{m} \text{tr}\ (X - G_j Y_j)' W_j (X - G_j Y_j), \tag{3}$$

The W_j are diagonal matrices with

$$w_i^j := w_{il(i,j)} = \sum_{l=1}^{k_j} g_{il}^j w_{il}^j. \tag{4}$$

In homogeneity analysis we minimize (3) 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

$$X\Lambda = W_\star^{-1} \sum_{j=1}^{m} W_j G_j Y_j, \tag{5}$$

$$Y_j = (G_j' W_j G_j)^{-1} G_j' W_j X, \tag{6}$$

where Λ is a symmetric matrix of Lagrange multipliers.

2 Loss Function

Now consider the closely related problem in which we do not require, as in homogeneity analysis, that

$$\hat{d}^j_{il(i,j)} = 0$$

but we impose the weaker condition that $\hat{d}^j_{il(i,j)}$ is less than or equal to all $\hat{d}^j_{il}$ in row i.

In homogeneity analysis the geometric interpretation of loss is that we want objects to coincide with the categories they score in (for all variables). 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 (for all variables). The plot of the the k_j categories of variable r defines k_j Voronoi regions. The Voronoi region of category ℓ is the polyhedral convex set of all points closer to category ℓ than to any other category of variable r. The loss function … vanishes if the x_i are all in the Voronoi regions of the categories they score in (for all variables). It is sufficient for contiguity that the stars in the star plot are disjoint ? Also that the convex hulls of the object points are disjoint ?

2.1 The Unconstrained Case

ALS

Minimizing … over the rows δ^r_i is a monotone regression for a simple tree order. This is easily handled by using Kruskal's primary approach to ties (Kruskal (1964a), Kruskal (1964b), De Leeuw (1977)).

In stage 2 we do one or more metric smacof iterations for given Δ_j to decrease the loss. These 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)).

Thus for stage two purposes the loss function is

$$\sigma(Z_1, \cdots, Z_m) = \sum_{j=1}^{m} \sum_{i=1}^{N_j} \sum_{j=1}^{N_j} w^r_{ij} (\delta^r_{ij} - d_{ij}(Z_j))^2,$$

with $N_j := n + k_j$ and

$$Z_j := \begin{matrix} n \\ k_j \end{matrix} \begin{bmatrix} \overset{p}{X} \\ Y_j \end{bmatrix}.$$

The weights in W_j are now zero for the two diagonal blocks.

$$g^r_i := \begin{matrix} n \\ k_j \end{matrix} \begin{bmatrix} e_i \\ 0 \end{bmatrix}.$$

$$h_\ell^r := \begin{matrix} n \\ k_j \end{matrix} \begin{bmatrix} 0 \\ e_\ell \end{bmatrix}$$

so that $x_i = Z_j' g_i^r$ and $y_\ell^r = Z_j' h_\ell^r$. It follows that

$$d^2(x_i, y_\ell^r) = \text{tr } Z_j' A_{i\ell}^r Z_j$$

where $A_{i\ell}^r$, of order $n + k_j$, is the rank-one, positive semi-definite, doubly centered matrix

$$A_{i\ell}^r := (g_i^r - h_\ell^r)(g_i^r - h_\ell^r)'$$

It further follows that

$$\sum_{i=1}^{n} \sum_{\ell=1}^{k_j} w_{il}^r d^2(x_i, y_\ell^r) = \text{tr } Z_j' V_j Z_j,$$

where

$$V_j := \sum_{i=1}^{n} \sum_{\ell=1}^{k_j} w_{il}^r A_{i\ell}^r,$$

so that V_j has the shape

$$V_j = \begin{matrix} \\ n \\ k_j \end{matrix} \begin{matrix} \begin{matrix} n & k_j \end{matrix} \\ \begin{bmatrix} \ddots & \square \\ \square & \ddots \end{bmatrix} \end{matrix}$$

with $-W_j$ in the off-diagonal block, while the two diagonal blocks are diagonal matrices with the row and column sums of W_j.

In a similar way we define $B_j(Z_j)$, with the same shape as V_j, and with

$$B(Z_j) = \sum_{i=1}^{n} \sum_{\ell=1}^{k_j} w_{i\ell}^r \frac{\delta_{i\ell}^r}{d_{i\ell}(Z_j)} A_{i\ell}^r$$

$$\sigma(Z_1, \cdots, Z_m) = \sum_{j=1}^{m} \{\eta_j^2 - 2\text{tr } Z_j' B(Z_j) Z_j + Z_j' V_j Z_j\}$$

with

$$\eta_j^2 := \sum_{i=1}^{N_j} \sum_{j=1}^{N_j} w_{ij}^r \{\delta_{ij}^r\}^2$$

Define the Guttman transforms of Z_j as

$$\tilde{Z}_j = V_j^+ B(Z_j) Z_j$$

Then

$$\sigma_j(Z) := \eta_j^2 - 2\text{tr } Z_j' B(Z_j) Z_j + Z_j' V_j Z_j = \eta_j^2 - \eta^2(\tilde{Z}_j) + \text{tr } (Z_j - \tilde{Z}_j)' V_j (Z_j - \tilde{Z}_j)$$

If $Z_j^{(k)}$ is Z_j at iteration k then

$$\text{tr } Z_j' B(Z_j) Z_j \geq \text{tr } Z_j' B(Z_j^{(k)}) Z_j^{(k)} = \text{tr } Z_j' V_j \tilde{Z}_j^{(k)}.$$

Thus

$$\sigma_j(Z) \leq \eta_j^2 - \eta^2(\tilde{Z}_j^{(k)}) + \text{tr } (Z_j - \tilde{Z}_j^{(k)})' V_j (Z_j - \tilde{Z}_j^{(k)})$$

It follows that in iteration $k+1$ we must minimize

$$\omega(Z_1, \cdots, Z_s) := \sum_{j=1}^m \text{tr } (Z_j - \tilde{Z}_j^{(k)})' V_j (Z_j - \tilde{Z}_j^{(k)})$$

to compute the $Z_j^{(k+1)}$.

Computation of the Guttman transform requires Moore-Penrose inverses of the matrices V_j, which are of order $n + k_j$ and extremely sparse. They can also be uncomfortably large because n can be large. It is much more memory-friendly to solve the partioned system

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

More explicitly we must minimize

$$\sum_{j=1}^m \text{tr } (X - \tilde{X}_j)' R_j (X - \tilde{X}_j) - 2 \sum_{j=1}^m \text{tr } (X - \tilde{X}_j)' W_j (Y_j - \tilde{Y}_j) + \sum_{j=1}^m \text{tr } (Y_j - \tilde{Y}_j)' C_j (Y_j - \tilde{Y}_j) \quad (7)$$

where $\tilde{X}_j$ and $\tilde{Y}_j$ are the two components of the Guttman transform $\tilde{Z}_j$ of the current Z_j.

First minimize over Y_j for given X. This gives

$$Y_j = \tilde{Y}_j - \{V_{22}^j\}^{-1} V_{21}^j (X - \tilde{X}_j)$$

$$X = \{V_{11}^\star\}^{-1} \sum_{j=1}^m \left\{ V_{11}^j \tilde{X}_j - V_{12}^j (Y_j - \tilde{Y}_j) \right\}$$

$$\sum_{j=1}^m V_{11}^j (X - \tilde{X}_j) + \sum_{j=1}^m V_{12}^j (Y_j - \tilde{Y}_j) = 0$$

Or we substitute … into … we find that we have to minimize

$$\sum_{j=1}^m \text{tr } (X - \tilde{X}_j)' (R_j - W_j C_j^{-1} W_j') (X - \tilde{X}_j)$$

over X. ## Missing Data

2.2 Normalization of X

Constraint

$$X'V_{11}^{\star}X = I$$

2.3 The Unweighted Case

In the unweighted case we have $V_{11}^r = k_j I$, $V_{22}^r = nI$, and $V_{12}^r = -E_{n\times k_j}$. Thus $V_{1|2}^r = k_j J_n$ and $V_{1|2}^{\star} = k_{\star} J_n$ with $J_n = I_n - n^{-1}E_{n\times n}$, the centering matrix of order $n.

The Guttman transform also simplifies in the unweighted case. The formulas were already given in Heiser and De Leeuw (1979).

2.4 Centroid Constraints on Y

$$Z_j = \begin{matrix} p \\ n \\ k_j \end{matrix} \begin{bmatrix} I \\ D_j^{-1}G_j' \end{bmatrix} X = H_j X$$

$$\sum_{j=1}^m Z_j'V_jZ_j = X'\{\sum_{j=1}^m H_j'V_jH_j\}X.$$

$$X'H_j'V_jH_jX = X'\{V_{11}^j + V_{12}^jD_j^{-1}G_j + G_jD_j^{-1}V_{22}^jD_j^{-1}G_j'\}X$$

2.5 Rank Constraints on Y

$$y_j^r = \alpha_j^r y_j$$

3 Note

In order to not have to deal with general inverses and multiplications of gigantic symmetric matrices with largely empty diagonal blocks we solve the system

$$\begin{bmatrix} W_j & -W \\ -W' & W_c \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} P \\ Q \end{bmatrix}$$

for X and Y, with the known right hand side

$$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} B_j & -B \\ -B' & B_c \end{bmatrix} \begin{bmatrix} X^{(k)} \\ Y^{(k)} \end{bmatrix}$$

$W_jX - WY = P$ or $X = W_j^{-1}(P + WY)$. Substitute in $W_cY - W'X = Q$ to get $W_cY - W'W_j^{-1}(P + WY) = Q$ or $(W_c - W'W_j^{-1}W)Y = Q + W'W_j^{-1}P$.

Note that $W_c - W'W_j^{-1}W$ is doubly-centered and $Q + W'W_j^{-1}P$ is column-centered.

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. 1977. “Correctness of Kruskal’s Algorithms for Monotone Regression with Ties.” *Psychometrika* 42: 141–44.

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. 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.