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 04, 2024

Abstract

smacofVO

Contents

1 Simultaneous Non-Metric Unfolding **3**
1.1 Homogeneity Analysis 3

2 Loss Function **5**
2.1 The Unconstrained Case 5
2.2 Missing Data 7
2.3 Normalization of X 7
2.4 The Unweighted Case 8
2.5 Centroid Constraints on Y 8
2.6 Rank Constraints on Y 8

3 Note **9**

References **10**

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

If we have data where the same individuals give preference judgments over multiple domains, or over the same domain on multiple occasions, or over the same domain under different experimental conditions, then we can use 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 individual and occasion there are different sets of transformations Δ_j and for each occasion there different matrices of column scores Y_j, but there is only a single matrix of row scores X.

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.

We give a somewhat non-standard introduction to homogeneity analysis here, to highlight the similarities with unfolding and the techniques we will present later on in this paper.

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

Homogeneity analysis makes a joint maps in p dimensions of individuals and categories (both represented as points) in such a way that category points are close to the points for the individuals in the category. And, vice versa, individuals 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 individual points coincide with their category points. But there are $s > 1$ indicator matrices, corresponding with s categorical variables, and the solution is a compromise trying to achieve homogeneity as well as possible for all variables simultaneously.

Let us use loss function (1) to captures loss of homogeneity in the sense discussed above. The sets Δ_i^r 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^r we have

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

where the $l(i,j)$ on the right is the index of the category of variable j that individual i is in. Using indicators we can write this as

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

with the weights now defined as

$$w_i^j := w_{il(i,j)}^j.$$

The other w_{il}^j do not matter and play no part in the optimization problem.

star plot

$$Y_r = (G_r' W_r G_r)^{-1} G_r' W_r X$$

$$\min_Y \sigma(X, Y_1, \cdots, Y_s) = \text{tr } X' \left\{ \sum_{r=1}^{s} \{W_r - W_r G_r (G_r' W_r G_r)^{-1} G_r' W_r\} \right\} X$$

$X' W_\star X = I$

2 Loss Function

2.1 The Unconstrained Case

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 merely require that $\hat{d}^j_{il(i,j)}$ is less than or equal to all $\hat{d}^j_{il}$ in row i. Formally

$$g^j_{il} = 1 \Rightarrow \hat{d}^j_{il} \leq \hat{d}^j_{iv} \qquad \forall v \neq \ell$$

In homogeneity analysis the geometric interpretation of loss is that we want individuals to coincide with the categories they score in (for all variables). The geometric interpretation of loss function … is that we want individuals 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_r categories of variable r defines k_r 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).

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_r 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

$$\sigma(Z_1, \cdots, Z_m) = \sum_{r=1}^{s} \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} w^r_{ij}(\delta^r_{ij} - d_{ij}(Z_r))^2,$$

with $N_r := n + k_r$ and

$$Z_r := \begin{matrix} n \\ k_r \end{matrix} \begin{bmatrix} \overset{p}{X} \\ Y_r \end{bmatrix}.$$

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

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

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

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

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

where $A_{i\ell}^r$, of order $n + k_r$, 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

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

where

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

so that V_r has the shape

$$V_r = \begin{matrix} \\ n \\ k_r \end{matrix} \begin{matrix} \begin{matrix} n & k_r \end{matrix} \\ \begin{bmatrix} \ddots & \square \\ \square & \ddots \end{bmatrix} \end{matrix}$$

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

In a similar way we define $B_r(Z_r)$, with the same shape as V_r, and with

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

$$\sigma(Z_1, \cdots, Z_m) = \sum_{r=1}^{s} \{\eta_r^2 - 2\text{tr } Z_r' B(Z_r) Z_r + Z_r' V_r Z_r\}$$

with

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

Define the Guttman transforms of Z_r as

$$\tilde{Z}_r = V_r^+ B(Z_r) Z_r$$

Then

$$\sigma_r(Z) := \eta_r^2 - 2\text{tr } Z_r' B(Z_r) Z_r + Z_r' V_r Z_r = \eta_r^2 - \eta^2(\tilde{Z}_r) + \text{tr } (Z_r - \tilde{Z}_r)' V_r (Z_r - \tilde{Z}_r)$$

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

$$\text{tr } Z_r'B(Z_r)Z_r \geq \text{tr } Z_r'B(Z_r^{(k)})Z_r^{(k)} = \text{tr } Z_r'V_r\tilde{Z}_r^{(k)}.$$

Thus

$$\sigma_r(Z) \leq \eta_r^2 - \eta^2(\tilde{Z}_r^{(k)}) + \text{tr } (Z_r - \tilde{Z}_r^{(k)})'V_r(Z_r - \tilde{Z}_r^{(k)})$$

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

$$\omega(Z_1, \cdots, Z_s) := \sum_{r=1}^{s} \text{tr } (Z_r - \tilde{Z}_r^{(k)})'V_r(Z_r - \tilde{Z}_r^{(k)})$$

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

More explicitly we must minimize

$$\sum_{j=1}^{m} \text{tr } (X - \tilde{X}_j)'V_{11}^j(X - \tilde{X}_j) + 2\sum_{j=1}^{m} \text{tr } (X - \tilde{X}_j)'V_{12}^j(Y_j - \tilde{Y}_j) + \sum_{j=1}^{m} \text{tr } (Y_j - \tilde{Y}_j)'V_{22}^j(Y_j - \tilde{Y}_j) \quad (2)$$

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

First minimize over Y_r for given X. This gives

$$V_{21}^j(X - \tilde{X}_j) + V_{22}^j(Y_j - \tilde{Y}_j) = 0$$

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

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

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

2.2 Missing Data

2.3 Normalization of X

Constraint either (weak)

$$\text{tr } X'V_{11}^\star X = 1$$

or (strong)

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

2.4 The Unweighted Case

In the unweighted case we have $V_{11}^r = k_r I$, $V_{22}^r = nI$, and $V_{12}^r = -E_{n\times k_r}$. Thus $V_{1|2}^r = k_r J_n$ 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).

2.5 Centroid Constraints on Y

$$Z_r = \begin{matrix} p \\ n \\ k_r \end{matrix} \begin{bmatrix} I \\ D_r^{-1}G_r' \end{bmatrix} X = H_r X$$

$$\sum_{j=1}^m Z_r' V_r Z_r = X'\{\sum_{j=1}^m H_r' V_r H_r\} X.$$

$$X' H_r' V_r H_r X = X'\{V_{11}^j + V_{12}^j D_j^{-1} G_j + G_j D_j^{-1} V_{22}^j D_j^{-1} G_j'\} X$$

2.6 Rank Constraints on Y

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

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

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

for X and Y, with the known right hand side

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

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

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

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.