HUDM5124: Introduction to Multidimensional Scaling, Clustering, and Related Methods

Session 6:

- Weighted (Individual Differences) MDS Model for 3-way Data.
- Algorithms

    SINDSCAL

    MDS by majorization (SMACOF)

WEIGHTED MDS

The main model we will discuss today, *weighted MDS* (WMDS), generalizes the Euclidean distance model so that several proximity matrices can be fit to a single underlying psychological space. WMDS incorporates a model to account for individual differences (between-subjects, or at least between-matrices) in how separate dimensions of the underlying psychological space are weighted. The basic idea is that each dimension may be weighted differently by different subjects.

Use of different dimension weights by different subjects means that each subject's individual psychological space is stretched or shrunk along each dimension, thus subjects' individual proximity matrices can differ from the group space systematically, and these systematic differences can be measured and modeled. For this reason, WMDS is often called *individual differences MDS.*

Weighted MDS: the problem

GIVEN: a set of m matrices Δ of proximities among n objects, that are assumed to be [linearly, monotonically] related to distances in a psychological space (with weights for dimensions allowed to vary across individuals or matrices)

ESTIMATE:

1) an N x R matrix X (= the configuration of N points in a geometric space of R dimensions), the "group space"
2) A set of nonnegative weights Wrp (r=1 to R, p=1 to m), indicating the weight or salience of dimension r to individual p.

Of course, the "correct" dimensionality must be determined as well.

The Weighted Euclidean Model

The simplest version of WMDS uses the following definition for weighted Euclidean distance. This is often referred to as the INDSCAL model (Carroll & Chang, 1970), although that terminology blurs the distinction between the model and the method used to fit it.

In this model, the distance between stimulus I and j (for the k-th subject) can be expressed as:

$$d_{ij(k)} = \sqrt{\sum_{r=1}^{R} w_{kr} \left| X_{ir} - X_{jr} \right|^2}$$

where the w_{kr} are weights for subject (or "source") *k* on each of the *r*=1 to *R* dimensions.

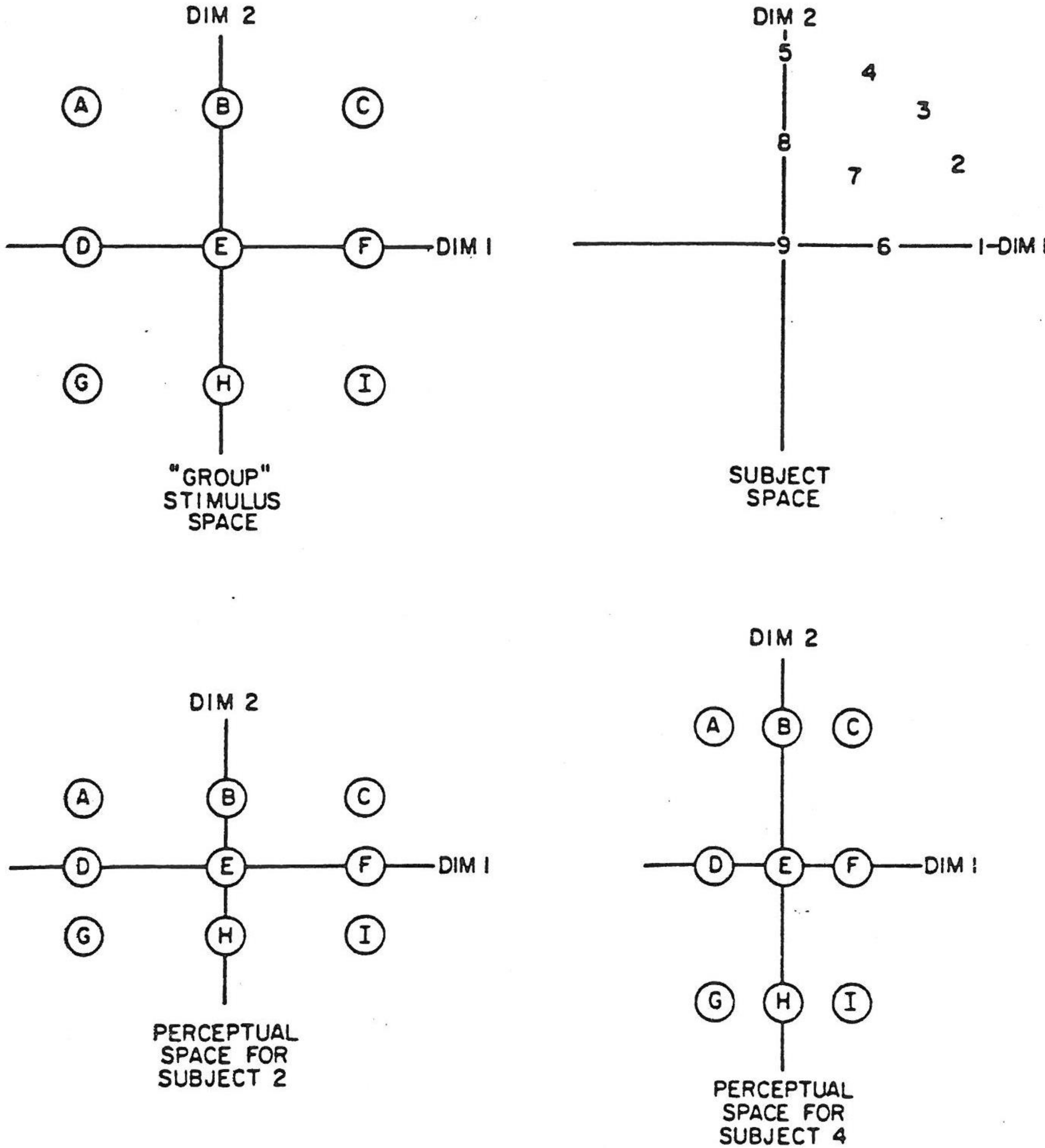
A generalization: the IDIOSCAL model

Carroll & Chang (1970) also proposed a more general weighted MDS model, which they termed IDIOSCAL. However, they did not provide an effective means of fitting/estimating it. In this model, individual subjects may rotate axes as well as rescale them.

$$d_{ij(k)} = \sqrt{\sum_{r=1}^{R} \sum_{r'=1}^{R} (X_{ir} - X_{jr}) c_{rr'(k)} (X_{ir'} - X_{jr'})}$$

Here, the $c_{rr'(k)}$ entries constitute an RxR matrix C (for each subject k) that can stretch or shrink axes (if it is diagonal), or rotate them (if it is a rotation matrix with non-zero off-diagonal elements).

An even more general model is **GEMscale** (Young, 1984, which adds another weighting matrix to accommodate Mahalanobis as well as Euclidean distances. GEM stands for “generalized Euclidean model”.



DIM 2
A
B
C
D
E
F
DIM I
G
H
I
"GROUP" STIMULUS SPACE
DIM 2
5
4
3
8
2
7
9
6
1
DIM I
SUBJECT SPACE
DIM 2
A
B
C
D
E
F
DIM I
G
H
I
PERCEPTUAL SPACE FOR SUBJECT 2
DIM 2
A
B
C
D
E
F
DIM I
G
H
I
PERCEPTUAL SPACE FOR SUBJECT 4

References:

- Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35,* 283-319. [algorithm underlying SINDSCAL program]
- Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika, 42*, 7-67. [ALSCAL algorithm]
- Arabie, P., Carroll, J. D., & DeSarbo, W. S. (1987). *Three-way scaling and clustering* (Sage University Papers series: Quantitative Applications in the Social Sciences, no. 07-65). Thousand Oaks: Sage
- Deleeuw, J. (1977). In R. Barra et al., Eds. *Recent Developments in Statistics.* North-Holland. Amsterdam. [algorithm underlying SMACOF program]
- Deleeuw, J. and Heiser, W. J. (1977). In J. C. Lingoe, Ed. *Geometrical Representatioms of Relational Data*. Mathesis Press. Ann Arbor, MI.

Availability of computer software for fitting the weighted MDS model:

program	Author	source
SINDSCAL	Carroll & Pruzansky	NETLIB
ALSCAL	Young	SPSS, SAS
SMACOF	DeLeeuw & Heiser (et al.)	SPSS(PROXSCAL), R

Some practical issues in weighted MDS:

Choosing the dimensionality. The problem of determining the dimensionality R of the solution space must be addressed. Often the dimensionality is selected *a priori* based on theory or *post-hoc* based on interpretability or on perception of an "elbow" in fit function (RSQ for INDSCAL; normalized raw stress for SMACOF).

How much data? For the metric models, having multiple matrices more than compensates for the extra parameters to be estimated.

Interpreting solutions by eye. In the INDSCAL model for weighted MDS, the orientation of the dimensions is NOT ARBITRARY (unlike 2-way MDS) – the stretching / shrinking of each dimension by individual subjects means that some optimal orientation of axes exists and is automatically identified.

Carroll & Chang (1970) INDSCAL algorithm to fit the weighted Euclidean model

Can be understood as an "n-way" generalization of Torgerson's metric MDS method

Generalize/re-express Torgerson's factorization problem:

$$b_{jk}^{(i)} = \sum_{t=1}^{r} y_{jt}^{(i)} y_{kt}^{(i)} = \sum_{t=1}^{r} w_{it} x_{jt} x_{kt} .$$

$$z_{ijk} \cong \sum_{t=1}^{r} w_{it} x_{jt}^{(L)} x_{kt}^{(R)}$$

Numerical technique is "alternating least-squares": first solve for W conditional on an estimate of X, then try to optimize X based on current estimate of W, etc.

SINDSCAL: version of INDSCAL that assumes symmetric data

SINDSCAL: version of INDSCAL that assumes symmetric data

SINDSCAL exists as a stand-alone FORTRAN program, available from NETLIB* and other web sites

*Netlib is a repository of software for scientific computing maintained by AT&T, Bell Laboratories, the University of Tennessee and Oak Ridge National Laboratory. Netlib comprises a large number of separate programs and libraries. Most of the code is written in Fortran, with some programs in other languages. See www.netlib.org

SINDSCAL is an effective algorithm. However, it fits the metric MDS model only.

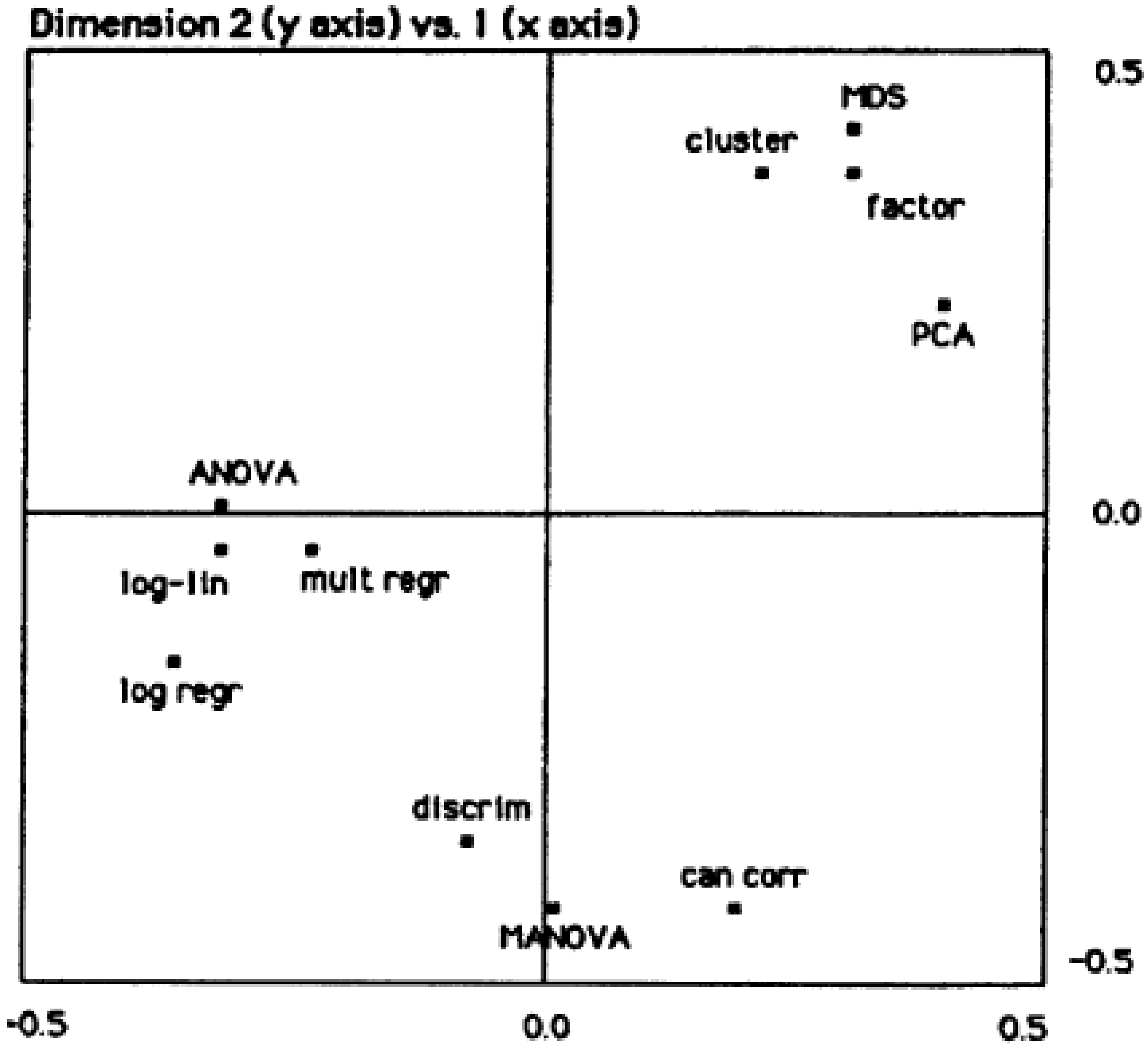
NOTE: R includes a “weighted CMDS” package that may or may not implement a similar algorithm (not confirmed).

An application (Corter & Carroll, 1990): rated substitutability of statistical techniques

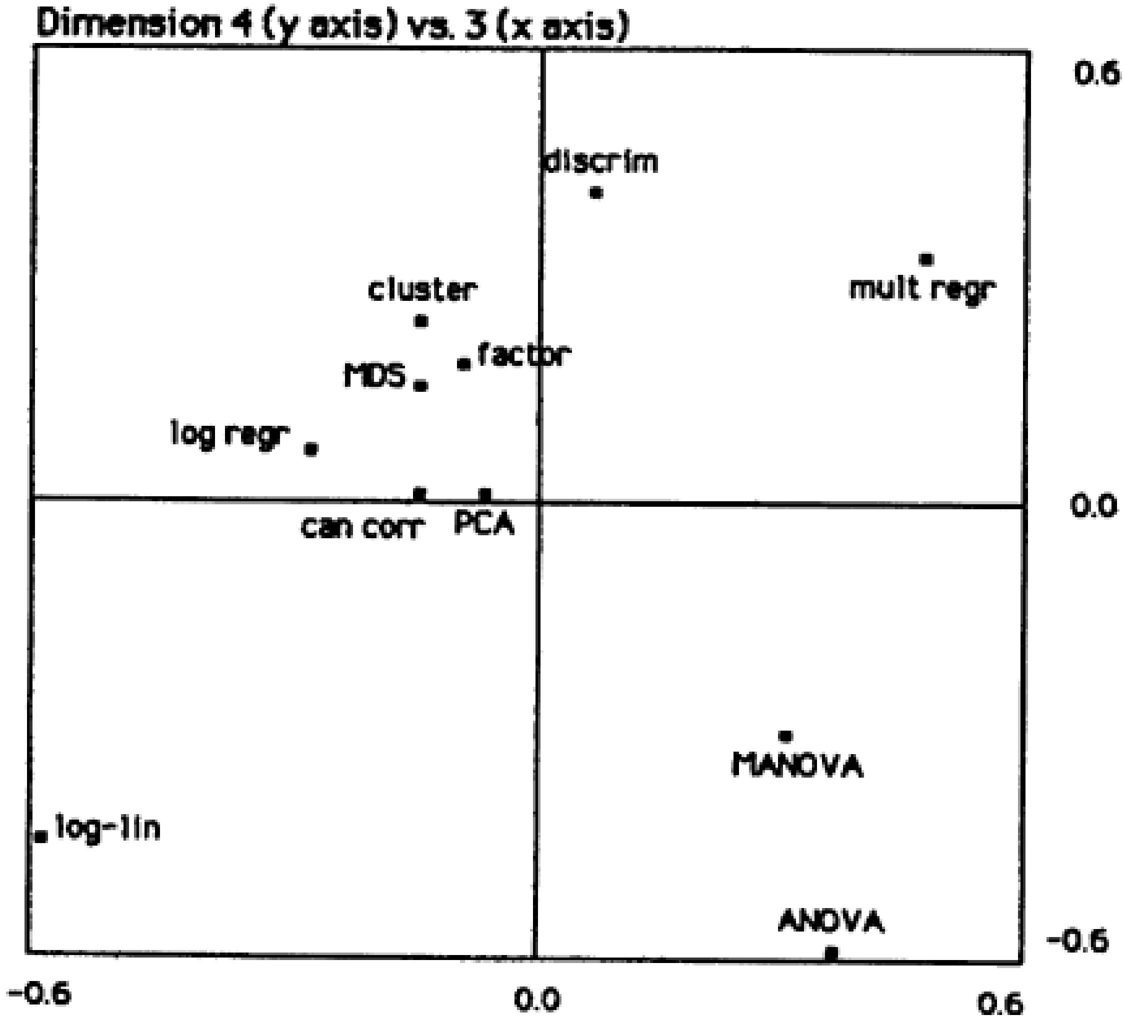
Table 1
Substitutability ratings for statistical techniques from three experts

	Expert A	Expert B	Expert C
MANOVA			
mult. reg.	2	7	9
PCA	2 2	4 2	3 1
factor	2 2 8	2 3 9	1 1 8
can. corr.	6 7 4 4	8 4 5 4	9 3 1 1
discrim.	8 6 3 2 3	8 6 3 4 8	9 5 1 1 9
cluster	3 4 7 8 5 3	1 2 5 7 3 4	1 1 5 5 1 1
MDS	3 4 8 8 5 3 8	1 2 8 8 3 2 6	1 1 5 7 1 1 7
ANOVA	5 3 1 1 2 1 1 1	7 7 1 1 4 3 2 1	9 9 1 1 2 7 1 1
log–lin.	6 8 1 2 4 5 3 2 8	5 2 1 1 4 2 1 1 6	2 2 1 1 3 3 1 1 2
log. reg.	6 8 1 3 5 8 1 2 8 8	4 7 2 2 4 6 3 1 6 8	6 6 1 1 6 6 1 1 7 8

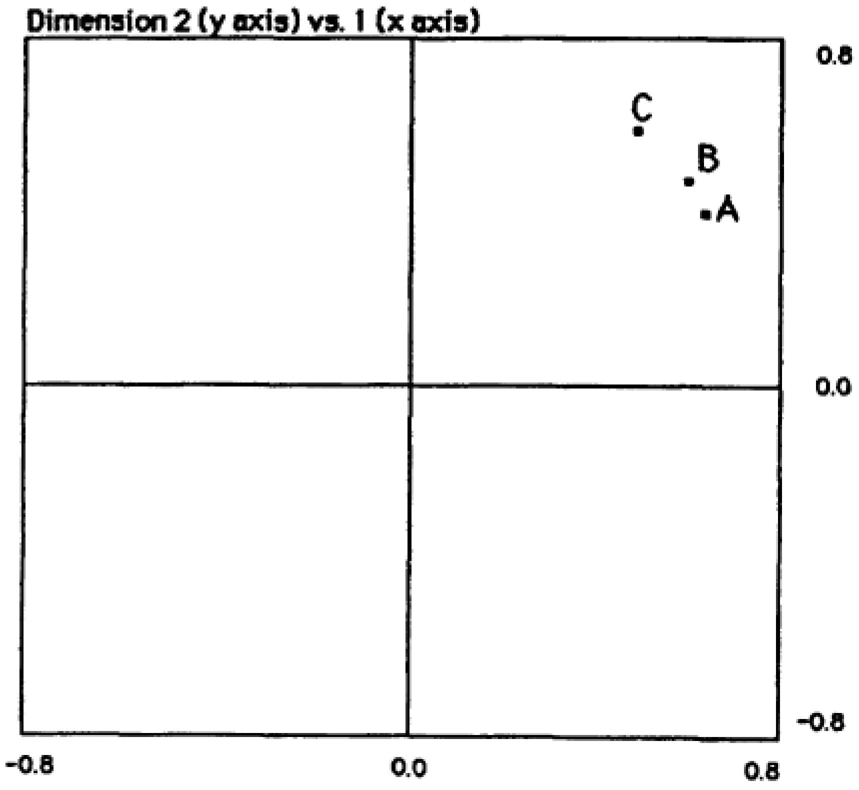
Corter & Carroll (cont): Dim. 1 vs. 2

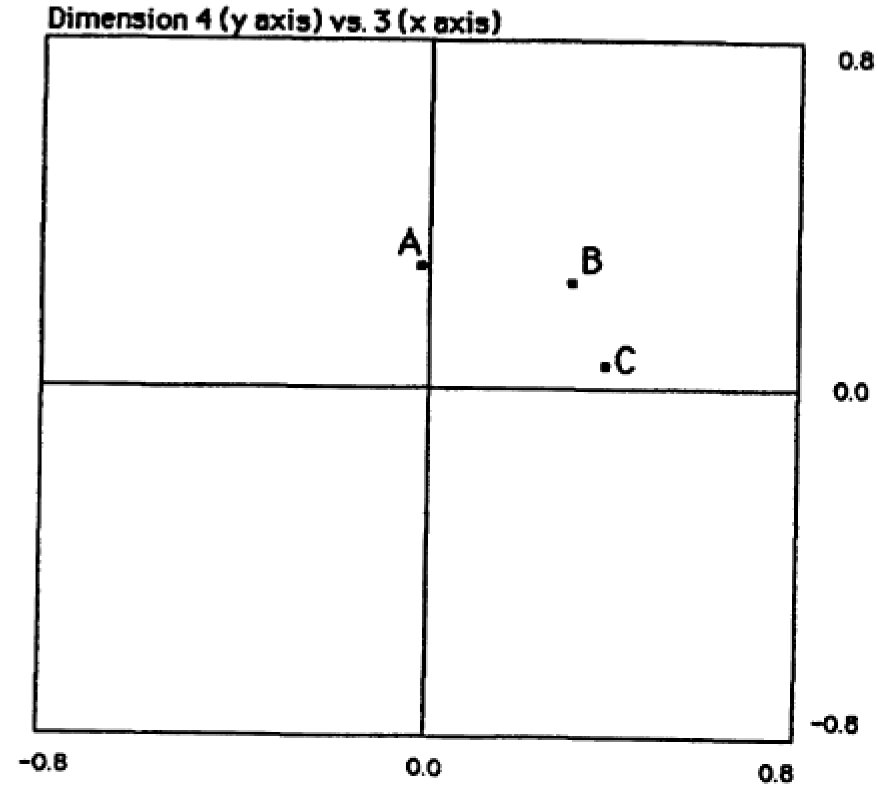


Corter & Carroll (cont): Dim. 3 vs. 4



Corter & Carroll (cont): subject weights space





Fitting MDS models via majorization (SMACOF)

SMACOF = Scaling by MAjorizing a COmplicated Function.

Refs: De Leeuw, J. (1977).

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. Journal of Classification, 5, 163–180.

Groenen, P. J. F., Mathar, R., & Heiser, W. J. (1995). The majorization approach to multidimensional scaling for Minkowski distances. Journal of Classification,12, 3-19.

→solve for X by minimizing raw stress: $\sigma^2(\mathbf{X}) = \sum_{i<j}^{n} w_{ij}(\delta_{ij} - d_{ij}(\mathbf{X}))^2 .$

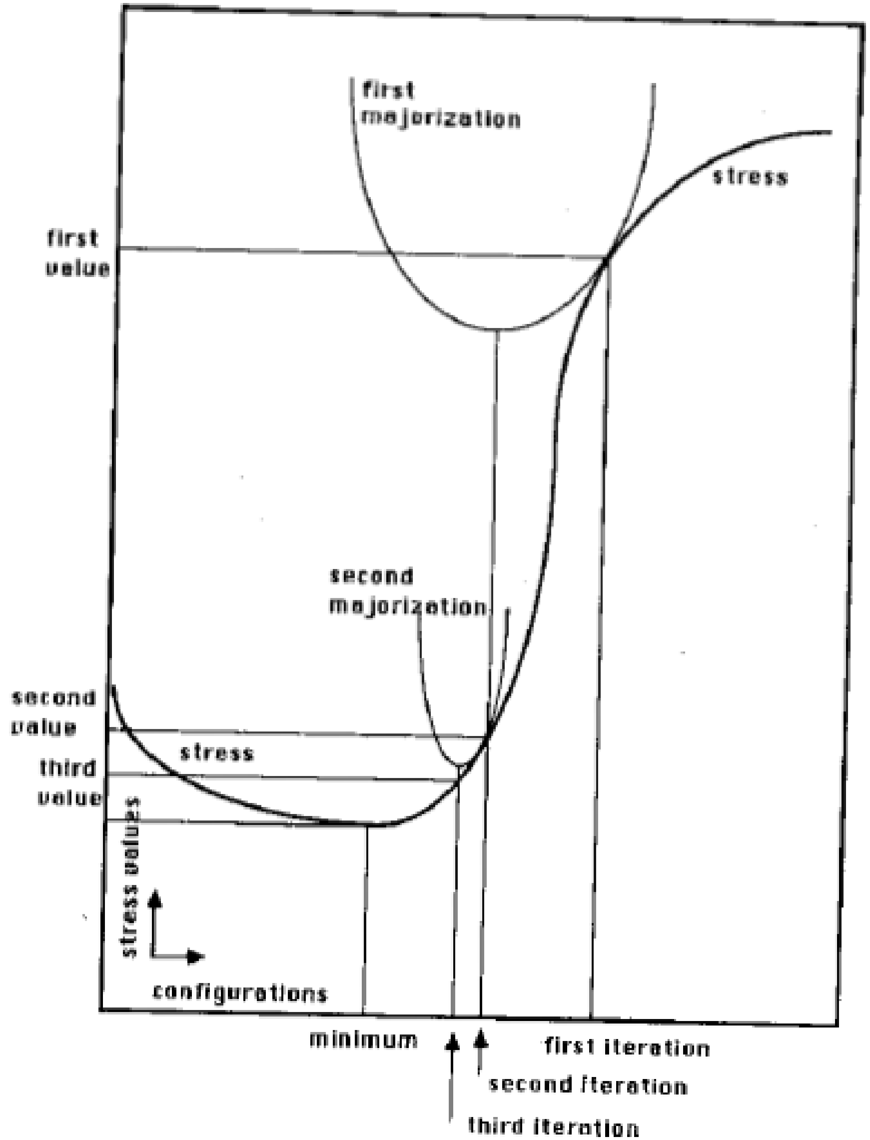
Rewrite as:

$$\sigma^2(\mathbf{X}) = \sum_{i<j}^{n} w_{ij}\,\delta_{ij}^2 + \sum_{i<j}^{n} w_{ij}\,d_{ij}^2(\mathbf{X}) - 2\sum_{i<j}^{n} w_{ij}\,\delta_{ij}\,d_{ij}(\mathbf{X})$$

$$= \eta_\delta^2 + \eta^2(\mathbf{X}) - 2\rho(\mathbf{X}) .$$

Majorization involves iterative minimization of this function.

Illustration of majorization



From de Leeuw (1988):

Figure 1. Three iterations of the majorization algorithm. We have drawn a section of the stress loss function, and two quadratic majorization functions. These touch the function at the current configuration, they are always above it, and their minimum provides the next configuration.

Minimizing a convex function by majorization

(from Groenen et al., 1995)

We describe the principle more formally. Let $\phi(\mathbf{X})$ be a real-valued function to be minimized over its domain $\mathcal{X}$. A function $\hat{\phi}(\mathbf{X},\mathbf{Y})$, $\mathbf{X},\mathbf{Y} \in \mathcal{X}$, with the properties

$$\phi(\mathbf{X}) \leq \hat{\phi}(\mathbf{X},\mathbf{Y}) \text{ and } \phi(\mathbf{X}) = \hat{\phi}(\mathbf{X},\mathbf{X}) \text{ for all } \mathbf{X},\mathbf{Y} \in \mathcal{X},$$

is called a majorizing function. Now, for fixed $\mathbf{Y}$ let $\mathbf{X}^+ = \arg\min_{\mathbf{X} \in \mathcal{X}} \hat{\phi}(\mathbf{X},\mathbf{Y})$ denote a minimum point. Then immediately we have the following chain of inequalities,

$$\phi(\mathbf{X}^+) \leq \hat{\phi}(\mathbf{X}^+,\mathbf{Y}) \leq \hat{\phi}(\mathbf{Y},\mathbf{Y}) = \phi(\mathbf{Y}), \tag{4}$$

which is named the *sandwich* inequality by De Leeuw (1992). The equality in (4) only occurs if $\mathbf{X}^+$ is also a stationary point of $\phi(\mathbf{X})$. The majorization algorithm can be summarized as

1. $\mathbf{Y} \leftarrow \mathbf{Y}_0$.
2. Find $\mathbf{X}^+$ for which $\hat{\phi}(\mathbf{X}^+,\mathbf{Y}) = \min_{\mathbf{X}} \hat{\phi}(\mathbf{X},\mathbf{Y})$.
3. If $\phi(\mathbf{Y}) - \phi(\mathbf{X}^+) < \varepsilon$ then stop. (ε a small positive constant.)
4. $\mathbf{Y} \leftarrow \mathbf{X}^+$ and go to 2.

SMACOF in SPSS: PROXSCAL

HANDOUTS:

Proxscal manual

EXAMPLE: kinship data (Andrade)

SPSS input file

SPSS output file

In R: INDSCAL via the smacof package

HANDOUTS:

The smacofIndDiff (a.k.a. indscal) routine in smacof

Example 1: kinship data

Example 2: wine tasting data (Mair, de Leeuw & Lienbacher)

Example 1: kinship data

```
# declare all prox matrices to be type "dist"
kin1<-as.dist(kin1)
kin2<-as.dist(kin2)
kin3<-as.dist(kin3)
kin4<-as.dist(kin4)
kin5<-as.dist(kin5)
kin6<-as.dist(kin6)
# now define input object "kinshipall" as a list of distance matrices:
kinshipall<-list(k1=kin1,k2=kin2,k3=kin3,k4=kin4,k5=kin5,k6=kin6)
kinshipall

# now fit the INDSCAL model to the data:
kin_ind<- indscal(kinshipall,type="interval",init="torgerson",verbose=TRUE)
# NOTE: can also specify "idioscal" model; “smacofIndDiff” will also fit indscal model
# plot the group stimulus space
plot(kin_ind$gspace,asp=1,pch=' ')
text(kin_ind$gspace,colnames(kinship))
# print out subject weights space:
kin_ind$cweights
```

Smacof in R: smacofIndDiff

3-Way SMACOF

SMACOF for individual differences:

- $k = 1, \ldots, K$ separate symmetric distance matrices.
- Data cube, or, in R: List.
- Classical approach: INDSCAL (Carrol & Chang, 1970).

Smacof in R: smacofIndDiff

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

5th Panhellenic Data Analysis Conference 2009-09-11

Smacof in R: smacofIndDiff

3-Way SMACOF

SMACOF for individual differences:

- $k = 1, \dots, K$ separate symmetric distance matrices.
- Data cube, or, in R: List.
- Classical approach: INDSCAL (Carrol & Chang, 1970).

Smacof in R: smacofIndDiff

Example 4: Wine Tasting

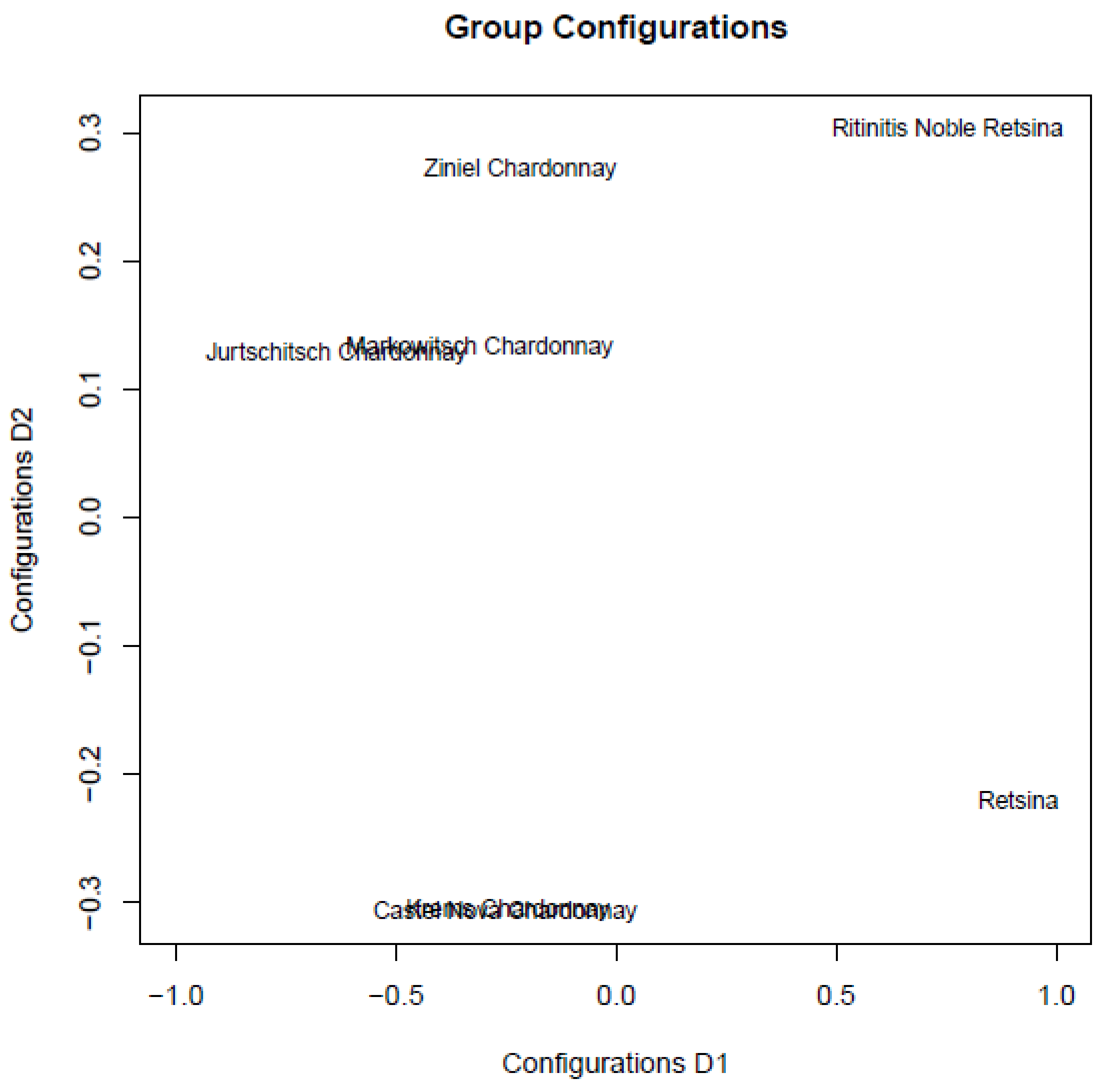
- Ziniel Chardonnay
- Markowitsch Chardonnay
- Krems Chardonnay
- Castel Nova Chardonnay
- Ritinitis Noble Retsina
- Retsina

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

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

```
R> plot(reswine, xlim = c(-1, 1))
```

Smacof in R: smacofIndDiff



Smacof in R: smacofIndDiff

Wine Tasting: Descriptives

	Price	Alcohol	Mean Rating
Jurtschitsch Chardonnay	14.99	13.00	2.00
Ziniel Chardonnay	7.00	12.00	2.60
Markowitsch Chardonnay	9.99	12.50	2.60
Ritinitis Noble Retsina	9.99	12.00	4.30
Retsina	2.99	11.50	4.60
Krems Chardonnay	5.99	12.50	2.70
Castel Nova Chardonnay	1.99	12.00	2.80

Smacof in R: smacofIndDiff

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.