

# MDS Extensions: Weighted MDS, Multidimensional Unfolding and the Vector Model for Unfolding

Measurement, Scaling, and Dimensional Analysis  
2019 ICPSR Summer Program  
Prof. Adam M. Enders

# Weighted MDS

- Sometimes referred to as the “weighted Euclidean” or “INDSCAL” (individual differences scaling) model
- Variant of standard, “two-way” MDS where more than one data set of object-pair dissimilarities is analyzed simultaneously
- WMDS takes “three-way” data, where the third “way” of the dataset is different subjects/measurements/etc.
- Objective: find lowest dimensional space that dissimilarities from several sources can be located in
- Output: “group configuration” and dimension weights that describe relative importance of dimensions for various subjects/measurements/etc.
  - ▶ Can “stretch” and “shrink” configuration by applying dimension weights to coordinates

# Weighted MDS

- The general WMDS model being fit is as follows:

$$d_{ijk}^2 = (x_i - x_j)\mathbf{W}_k(x_i - x_j)'$$

- $d_{ijk}^2$  represents the squared distance between object  $i$  and  $j$ , as modified by weight  $w$  for subject  $k$
- $x_i$  and  $x_j$  are stimulus coordinates for objects  $i$  and  $j$
- In classical MDS, the weight matrix would simply be an identity matrix with 1's on the diagonal and 0's on the off-diagonal since there are no weights
- $k$  refers to the “third way” of the data
- The weight matrix,  $\mathbf{W}_k$ , is a square, symmetric, (and, in this case) diagonal matrix of weights for each respondent type

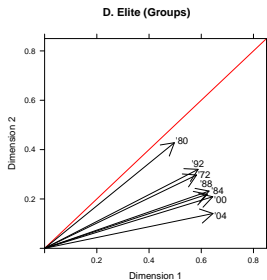
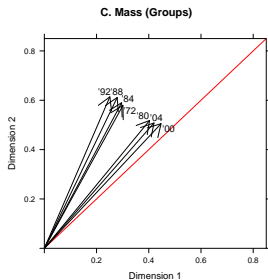
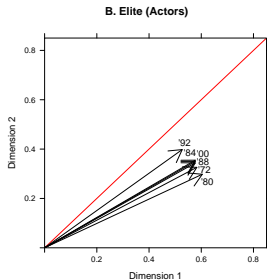
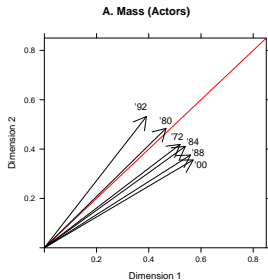
# Weighted MDS

- Applications:
  - ▶ Compare one group to another (e.g., differences in psychological structure)
  - ▶ Compare treatment groups to control (e.g., does frame/prime matter?)
  - ▶ Look for change at different time points
- Group configuration should make sense and fit should be “good” according to normal criteria (e.g., Stress, SPP, permutation test)
- Researchers are usually interested in the dimension weights
  - ▶ Can do statistical tests if differences in weight across groups/subjects
  - ▶ Can use weights as variables in the subsequent analyses

# Weighted MDS

**See “Weighted MDS Examples.r” script for example**

# WMDS Weights: Plotting



# WMDS Weights: Testing Differences

		Sum of Squares	df	Mean Square	F	P-value
Political Actor	Between	0.028	1	0.028	0.084	0.778
	Within	3.370	10	0.337		
	Total	3.398	11			
Political Group	Between	0.497	1	0.497	1.667	0.221
	Within	3.575	12	0.298		
	Total	4.072	13			

- Can do an ANAVA – analysis of angular separation – to test differences in angular separation across groups
- In this case, differences are not statistically significant

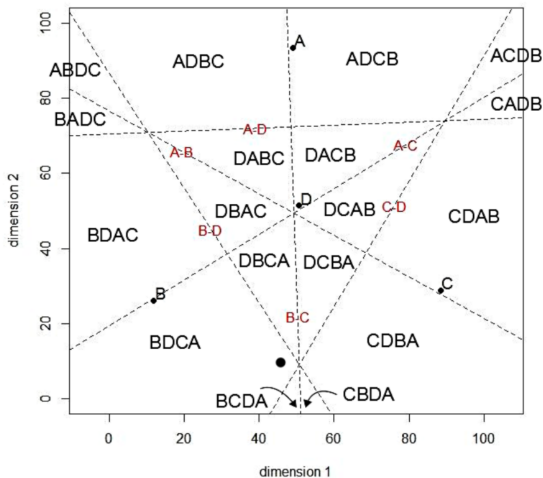
# Refresher on Unfolding

- Scaling model that assumes “single-peaked” item response functions, rather than monotonic ones, like with the cumulative scaling model
- Represents **proximities** between the  $n$  rows and  $k$  columns of a **rectangular** data matrix as distances between points along a single continuum
- Objectives:
  - ▶ Represent row and column objects in some low-dimensional space (two dimensions, ideally!)
  - ▶ Proximity between row and each column object should represent, to the best extent possible, the preferences/(dis)similarities from the original data matrix



# Refresher on Unfolding

Imagine a napkin!



# Multidimensional Unfolding

- Very similar to classical multidimensional scaling
- Major difference is that we need distances between row and column objects that are not the same
- Can also be estimated using the SMACOF algorithm in the “smacof” package in R using the “smacofRect” function
- Minimizes a loss function just like classical MDS

$$\sum_{i=1}^n \sum_{j=1}^q (d_{ij} - \delta_{ij})^2$$

- This looks just like Stress, except our error (differences between distances and dissimilarities) is summed over all pairs of  $n$  and  $q$ , since rows and columns are different

## Example: Jacoby 2014, cont'd

**Δ** matrix for a multidimensional unfolding analysis:

$$D_{n \times q} = \begin{bmatrix} \sum_{k=1}^s (x_{1k} - z_{1k})^2 & \sum_{k=1}^s (x_{1k} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{1k} - z_{qk})^2 \\ \sum_{k=1}^s (x_{2k} - z_{1k})^2 & \sum_{k=1}^s (x_{2k} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{2k} - z_{qk})^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{k=1}^s (x_{nk} - z_{1k})^2 & \sum_{k=1}^s (x_{nk} - z_{2k})^2 & \cdot & \cdot & \cdot & \sum_{k=1}^s (x_{nk} - z_{qk})^2 \end{bmatrix}$$

Software computes this for you – can just submit a rectangular matrix of proximity data

# Other Options for Estimation

1. Optimal Classification, developed by Poole (2000, 2005)
  - ▶ R package called “oc”
  - ▶ Can estimate unidimensional model, nonparametric
  - ▶ Only dichotomous data, programmed in language of legislators/votes
2. Ordinal Optimal Classification, Hare et al. (2018)
  - ▶ Nonparametric, ordinal data, can estimate unidimensional model
3. Smacof, developed by De Leeuw and colleagues (lots of papers)
  - ▶ R package of same name, ordinal and interval input data
  - ▶ Won't fit unidimensional models, no dichotomous data
  - ▶ In a pinch, could fit 2-dimensional model to unidimensional data and just use first dimension coordinates

# Multidimensional Unfolding

**See “Nonmetric Multidimensional Unfolding.r” script for example**

# Moving to the Vector Model of Unfolding...

- Think back to the summated rating model
- In that case, we collapsed across the columns of an  $n$  by  $k$  data matrix to create a unidimensional measure of some latent construct
- In doing this, we made several assumptions, one of which is that the items are “parallel measures” – they measure the same thing in roughly the same way
- What happens with a similar dataset, but the full profile of the rows across columns matters?
- Example: ranking a set of stimuli according to importance or attractiveness
  - ▶ In this case, we want to preserve the data within rows/across columns
  - ▶ Rankings where each cell value is a rank number are not parallel measures
  - ▶ Summing across the columns would result in scale scores identical for all rows

# The Vector Model of Unfolding

- So far, we've represented row objects as points in variable space (everything up to SVD)
- Also represented variables as both points (MDS, cumulative scaling, unfolding) and vectors (factor analysis, biplot)
- The Vector Model represents row objects as vectors emanative from the centroid, and column stimuli as points
  - ▶ Sometimes referred to as MDPREF, for MultiDimensional PReference Scaling
- Such a representation of the data allows for geometric interpretations of the relationships between pairs of row objects and between row objects and column objects
- The column points and row vectors are arranged so that *to the greatest extent possible*, row vectors point toward higher ranked column points, and away from lower ranked column objects

## Example

Hypothetical dataset with two individuals (row objects) and three generic stimuli (column objects)

Respondents asked to rank order the stimuli, where larger numerical values correspond to more preferred stimuli

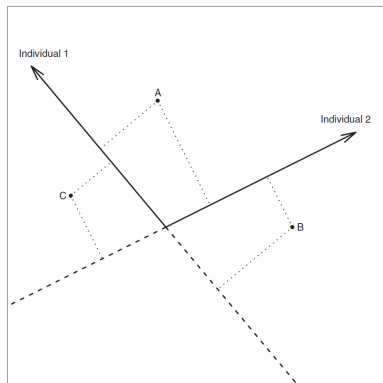
Stimuli:			
	A	B	C
Individual 1:	3	1	2
Individual 2:	2	3	1



## Example, cont'd

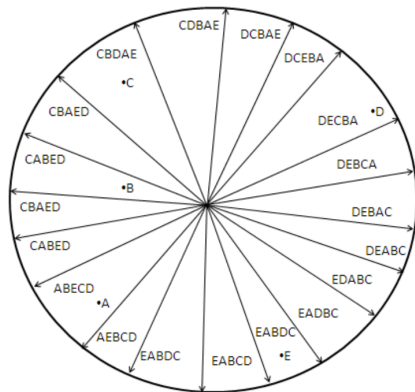
Objective is to locate column objects as points and row objects as vectors in a shared space

Drawing a perpendicular line from the column points to the row vectors should reproduce, as perfectly as possible, the original rank ordering in the dataset



## Example, cont'd

Expanded to more column and row objects:



Each vector reproduces the preferences/rank orders from the original data matrix

# The Model

- Assume some matrix  $\mathbf{X}$  contains  $n$  observations' scores for  $k$  stimuli
- Let the stimulus set,  $S_1, S_2, \dots, S_k$ , be represented by  $k$  points in  $m$ -dimensional space
- Point coordinates are collected into the  $k \times m$  matrix,  $\Phi$
- The observations,  $O_1, O_2, \dots, O_n$ , are represented by  $n$  unit-length vectors in the same  $m$ -space
- All of these vectors emanate from the origin, and their terminal points are collected into the  $n \times m$  matrix,  $\Gamma$

# The Model, cont'd

- Algebraically, the vector  $\gamma_i$  contains the terminal points of observation  $i$ 's vector within the  $m$ -space
- The projections from the stimulus points onto  $i$ 's vector are obtained by taking the scalar products of the stimulus points and the observation vector:

$$\hat{x}_i = \Phi \gamma_i$$

- We assume the dimensionality,  $m$ , should be relatively small relative to the dimensionality of the original data matrix
- If that's true, then we need to introduce an error term:

$$x_i = f(\hat{x}_i) + \epsilon_i = f(\Phi \gamma_i) + \epsilon_i$$

- $\epsilon_i$  is an error vector

# The Model, cont'd

- If the data are assumed to be interval level:

$$x_i = \alpha + \beta \hat{x}_i + \epsilon_i$$

- If the data are ordinal level:

$$x_i = f(\hat{x}_i) + \epsilon_i$$

- Here,  $f$  is a monotonic function

# Estimation: Metric Model (Interval/Ratio Data)

- First, we're going to consider how to estimate a metric model, which assumes interval or ratio level data
- Estimation is very easy: basically just a singular value decomposition
- Steps:
  1. Standardize rows of  $\mathbf{X}$  to 0 mean and unit variance,  $\mathbf{X}_{std}$
  2. Perform SVD of  $\mathbf{X}_{std} = \mathbf{UDV}'$ 
    - $\mathbf{U}$  in an  $n$  by  $q$  matrix of left singular vectors
    - $\mathbf{D}$  is the  $q$ -order diagonal matrix of singular values
    - $\mathbf{V}$  is the  $k$  by  $q$  matrix of right singular vectors
  3. Determine appropriate dimensionality,  $m$ , by considering model fit in a number of dimensions
    - $R^2 = tr(\mathbf{D}_m^2)/tr(\mathbf{D}^2)$ , where  $tr$  is the matrix trace, or sum of diagonal elements
  4. After determining  $m$ ,  $\Phi = \mathbf{V}_m$ , and  $\Gamma = \mathbf{U}_m \mathbf{D}_m$

## Estimation: Metric Model (Interval/Ratio Data)

**See “Vector Model.r” script for example**

## Estimation: Nonmetric Model (Ordinal Data)

- To do a estimate a nonmetric vector model, we just need to iteratively optimally scale the row and column coordinates to be monotonic functions of the original dataset
- Just like we did the first week with ALSOS
- Steps:
  1. Specify  $m$ , the dimensionality of the space, initialize  $R^2$  to zero, and initialize  $\mathbf{X}^*$  by setting it equal to the original  $\mathbf{X}$  matrix. Standardize within rows of  $\mathbf{X}^*$  to obtain  $\mathbf{X}_{std}^*$
  2. Perform SVD on  $\mathbf{X}_{std}^*$  to obtain  $\mathbf{U}_m$ ,  $\mathbf{D}_m$ , and  $\mathbf{V}_m$ . Use these matrices to calculate current estimates of  $\hat{\mathbf{X}}$  and  $R^2$
  3. If the current  $R^2$  is larger than the previous value, then continue. If  $R^2$  has not changed from the previous iteration, go to step 6



# Estimation: Nonmetric Model (Ordinal Data)

- Steps, cont'd:
  4. For  $i = 1, 2, \dots, n$ , use Kruskal's monotonic regression (1964) to find a new estimate of  $\hat{x}_i^*$  that are maximally correlated with the current model-based predicted values but always weakly monotonic to the original data
  5. Return to step 2 and carry out another iteration of the estimation procedure on the new version of  $\mathbf{X}^*$  that was obtained in step 4
  6. When  $R^2$  converges, construct the  $\Phi$  and  $\Gamma$  matrices from the singular vectors and values, and use the final  $R^2$  as the goodness-of-fit
- Row and column coordinates are calculated exactly as they were before, except the model will likely fit better due to the weaker assumption of (weak) monotonicity

## Estimation: Nonmetric Model (Ordinal Data)

**See “Vector Model.r” script for example**

## Example: Jacoby 2014

Respondents were confronted with “triads” of the following issues, from which rankings were constructed

**TABLE 1. Values and Definitions Shown to Respondents in Post-Election Wave of the 2006 Cooperative Congressional Election Study. In the Actual Administration of the Survey, the Order of the Values was Randomized Across Respondents**

---

FREEDOM, that is the widest liberty possible for everyone to act and think as they consider most appropriate.

EQUALITY, that is making sure that everyone has the same chance to get ahead in life.

ECONOMIC SECURITY, that is making sure that everyone has a steady job, a decent income, and a reasonable standard of living.

MORALITY, that is people living according to the rules that most people agree constitute decent human behavior.

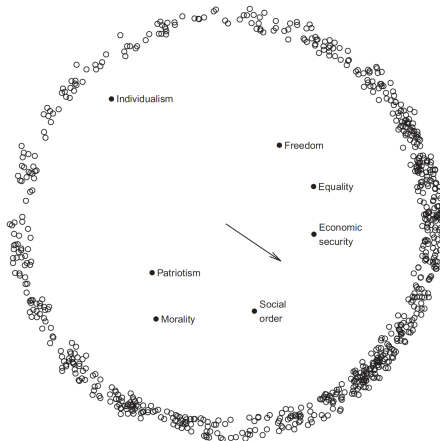
INDIVIDUALISM, that is everyone getting ahead in life on their own, without extra help from government or other groups.

SOCIAL ORDER, that is being able to live without fear, in a safe, peaceful society where the laws are respected and enforced.

PATRIOTISM, that is looking beyond our own personal interests and doing things that honor, respect, and protect our nation as a whole.

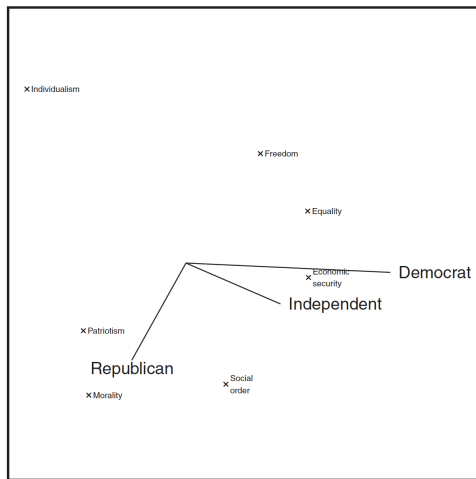
# Example: Jacoby 2014

Vector from dataset of rank-ordered value preferences



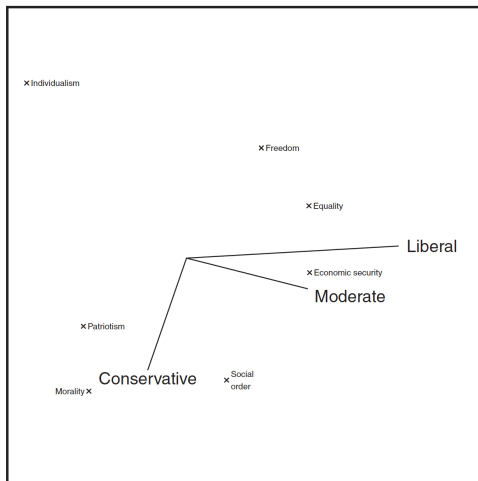
## Example: Jacoby 2014, cont'd

Plotting average vectors for partisan self-identifiers



## Example: Jacoby 2014, cont'd

Plotting average vectors for ideological self-identifiers



# Vectors and Angles

- Angles between vectors have same interpretation as with the biplot or factor analysis
- Correlation between vectors = cosine of the angle between the vectors:

$$\cos\theta_{xy} = \frac{x \cdot y}{|x||y|}$$

- The length of  $x = |x| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \dots + x_m^2}$
- Can use the vector coordinations to calculate average vector, or average vector for groups
  - ▶ Direction of mean vectors will correspond to average preference profile
  - ▶ Length of average vector will correspond to homogeneity of group preferences, where longer mean vectors denote greater homogeneity of preferences
- Can use Analysis of Angular Variation (ANAVA), the vector variant of ANOVA, to test whether group vectors are significantly different

# Power of the Vector Model

- We reduced dimensionality
- Made a picture that is fairly easy to interpret
- Used some basic geometry to visually represent:
  - ▶ The distribution of preferences
  - ▶ Average preferences of particular groups, and within group variance of those preferences
  - ▶ Correspondence (correlation) between preferences of individuals and groups
- Created a parsimonious measure of individual preferences that can be used in subsequent statistical analyses



## Example: Jacoby 2014, cont'd

Can use individual vector positions along the unit circle to do a “circular regression”

Conceptualize vector positions as angular deviations from some arbitrary, fixed vector position (say, 12:00)

**TABLE 3. Circular Regression Showing the Effects of Social and Political Characteristics on Individual Preference Vector Orientations**

	MLE Coefficient	Standard Error	Observed Probability
Respondent's Age	-0.008	0.002	0.000
Some College	0.148	0.051	0.002
College Graduate	0.429	0.065	0.000
Religious Commitment	-0.020	0.006	0.000
Party Identification	-0.109	0.016	0.000
Lib-Con Ideology	-0.140	0.026	0.000
Intercept ( $\mu$ )	-0.951	0.044	0.000

# Other MDS Models

- The IDIOSCAL model (variant of WMDS)
  - ▶ WMDS where matrix is not constrained to be diagonal (i.e., merely stretching and shrinking the group configuration)
  - ▶ This allows for unique dimensions for each individual/country/whatever the individual dissimilarities matrices correspond to
- MDS with External Constraints (variant of “Confirmatory MDS”)
  - ▶ Restrict objects to certain regions of the MDS space
  - ▶ Perhaps some points should cluster, and others not
- MDS with Internal Constraints (variant of “Confirmatory MDS”)
  - ▶ Constrain the point configuration to take a particular shape (geometric restrictions)
  - ▶ Most common is spherical MDS where points are constrained to align along a circle, sphere, or hypersphere