Multivariate techniques for sorting data: DiSTATIS and Discriminant DiSTATIS

Kriegsman, M.A.[1*], Raman, R.[1], Tillman, B.[2], Dowling, W.J.[1], & Abdi, H.[1]

[1]The University of Texas at Dallas, Richardson, TX, USA
[2]Lyon Neuroscience Research Center, Lyon, France
*Michael.Kriegsman@utdallas.edu

Introduction

Sorting:

- In a sorting task: assessors (participants) sort stimuli by similarity
  - Free vs constrained: Assessors may or may not be told the number of stimulus categories
  - Verbal descriptions of stimuli can also be analyzed, and can enrich interpretation[5]
- Fields: consumer preference (food, beer, wine, textile, perfume)[6,7] sensory evaluation (smell, taste, sound, touch), cognition (music)
- Applications: R&D, quality control, marketing[7]
- Advantages of sorting:
  - Requires minimal training. Amateurs and experts often give similar results[5,6], though similarity between amateurs and experts may differ by stimulus type[7]
  - Relatively easy, fast, and not fatiguing, even for many stimuli[5]
  - Does not require *a priori* selection of attributes/categories
- Disadvantages of sorting:
  - May be less accurate than ratings (a.k.a., profiling)[5,6]
- Analysis
  - Data: Square dissimilarity matrices
  - Analyzed often by MDS. MDS maps *overall* perceived similarity of stimuli
    - DiSTATIS reveals *individuals'* perceived similarity of stimuli
- Other potential applications of (Di)DiSTATIS: brain data (functional and/or structural connectivity)

References:

1. Abdi, H. (2007). Metric multidimensional scaling: Analyzing distance matrices. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 598–605). Thousand Oaks (CA): Sage.
2. Abdi, H., & Williams, L. J. (2010a). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(August), 433–459.
3. Abdi, H., & Williams, L. J. (2010b, June 23). Barycentric Discriminant Analysis. In *Encyclopedia of research design*.
4. Abdi, H., Williams, L. J., Valentin, D., & Bennani-Dosse, M. (2012). STATIS and DISTATIS: Optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*, 124–167.
5. Cartier, R., Rytz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., & Martin, N. (2006). Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, *17*(7-8), 562–571.
6. Chollet, S., Lelièvre, M., Abdi, H., & Valentin, D. (2011). Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask. *Food Quality and Preference*, *22*(6), 507–520.
7. Chollet, S., Valentin, D., & Abdi, H. (2014). Free Sorting Task. In P. V. Tomasco & G. Ares (Eds.), *Novel Techniques in Sensory Characterization and Consumer Profiling* (pp. 207–227). Boca Raton: Taylor and Francis.

(Di)DiSTATIS: PCA + dissimilarity + multitable (+ discriminant)

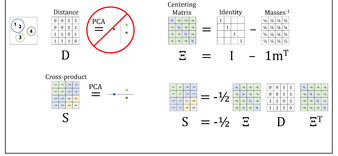
PCA (Principal Component Analysis):

- The core of these multivariate methods
- Eigen-decomposition (square psd matrices)
- Singular Value Decomposition (rectangular matrices)

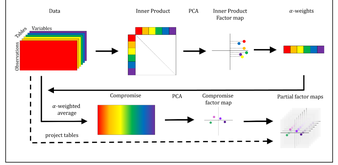
Eigen-decomposition: $X = U V U^T$; $U V^{\frac{1}{2}} =$ 

Singular Value Decomposition: $X = P \Delta Q^T$; $P \Delta =$ 

MDS (Metric Multi-Dimensional Scaling)
- Dissimilarity

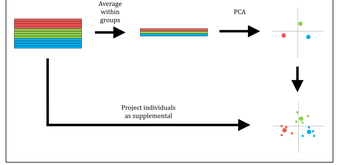


STATIS (*A Long French Name)
- Multitable

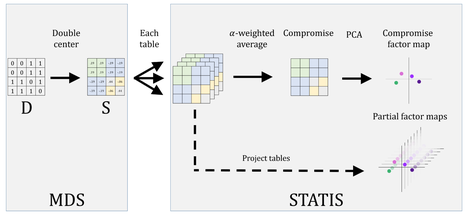


*Structuration des Tableaux ´a Trois Indices de la Statistique
Roughly translated: Structuring Three-Way Statistical Tables

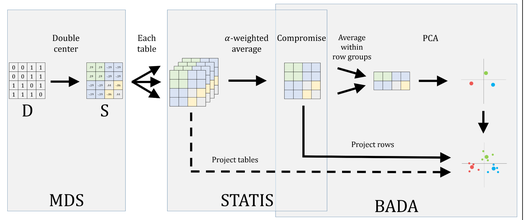
BADA (Barycentric Discriminant Analysis)
- Discriminant



DiSTATIS = MDS + STATIS



DiDiSTATIS = MDS + STATIS + BADA



An Example in Music Cognition

Experimental question:

- Do assessors perceive the stylistic differences between 3 composers, Bach, Beethoven, and Mozart?

Experimental Designs:

- Constrained sort into 3
- Unbeknownst to assessors, 3 composers
- Two Experiments: (1) MIDI sound clips, (2) recordings (by 4 pianists: Richter, Arrau, Pires, Barenboim)

Results Summary / Discussion:

- Better discrimination for MIDI > Recordings, DiDiSTATIS > DiSTATIS.
- Across all analyses: Beethoven opposed Mozart
- Recordings showed: Barenboim was Beethoven-like, Richter was Mozart-like
- No main effect of experience, except on recordings, where ↑ experience = ↑ sorting variability
  - Only some songs within each category were consistently sorted
- DiSTATIS vs DiDiSTATIS: DiDiSTATIS had ↑ %var on Component 1, but only a modest ↑ in R^2