Workshop in Descriptive Multivariate Methods in Social Sciences: Simple and Multiple Correspondence Analysis

Organised by Martyn Egan & Stefano Palestini

Lecturer: Johs. Hjellbrekke, University of Bergen, Norway.

Monitoring professor: Fabrizio Bernardi

May 17th-18th, 2012

Please register with Päivi.Kontinen@eui.eu

Course Description: Simple and multiple correspondence analysis

In the social sciences, multiple correspondence analysis (MCA) is a statistical technique that first and foremost has become known through the work of the late Pierre Bourdieu (1930-2002), in particular "Distinction" (Bourdieu 1984), "Homo Academicus" (Bourdieu 1988) and "The State Nobility" (Bourdieu 1996).

As a counterpart to principal component analysis (PCA), a geometric technique for the analysis of metric variables, MCA is a geometric technique for the analysis of categorical or categorized variables. Originating in the early 1960s and the French statistician Jean-Paul Benzécri's work in mathematical linguistics, MCA represents and models data sets as clouds of points in a multidimensional Euclidean space. The interpretation of the data is based on these clouds of points. By combining MCA with inferential techniques and variance analysis, we arrive at an integrated framework of interpretation that also is known under the name of Geometric Data Analysis (GDA).

In a combination of lectures and laboratory excercises, this course will introduce students to the fundamental properties, procedures and rules of interpretation of the most commonly used forms of correspondence analysis, i.e. simple correspondence analysis (CA) and MCA, and also to the most commonly used software.

Particular attention will be paid to how MCA can be used in the construction of social spaces.

The course starts with an historical introduction to Benzécri's work on contingency tables, and to the key ideas and basic properties in geometric data analysis. A first

explanation of the procedures, the key concepts and the fundamental rules of interpretation will be given through a simple correspondence analysis (CA) of a standard contingency table.

Thereafter, and using Pierre Bourdieu's work as a point of reference, we will go through the generalisation from CA to MCA by analysing an Individuals x Variables table. Particular emphasis will be put on the definition of distances between individuals, of distances between categories or modalities, the fundamental rules for the interpretation of axes in MCA, on how MCA can be integrated with variance analysis, and also on more general guidelines and coding principles.

We will then proceed to the more detailed exploration of the cloud of individuals, the introduction of supplementary variables, the use of concentration ellipses and how MCA also can be used in a confirmatory or explanatory mode by introducing variables as structuring factors in the constructed space. Tools for statistical inference, i.e. confidence ellipses around mean modality points in factorial planes, will also be presented, as will also the integration of MCA and hierarchical clustering as a way of further validation of the results.

Lectures and reading list:

Day 1: May 17st 2011

Lecture 1:

1. **General Introdution. Correspondence Analysis in the Social Sciences. Some Basic Properties in CA**

The BMS: (Karl M. van Meter, Marie-Ange Schiltz, Philippe Cibois and Lisa Mounier) (1994). "Correspondence Analysis: A History and French Sociological Perspective." In Greenacre, Michael J. & Blasius, Jörg, eds., *Correspondence Analysis in the Social Sciences: Recent Developments and Applications.* San Diego: Academic Press. Pp.128-138.

Le Roux, Brigitte & Henry Rouanet (2004). *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis.* Dordrecht: Kluwer Academic Publishers. Chapter 1: Overview of Geometric Data Analysis. Pp.1-22

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple Correspondence Analysis.* Series: Quantitative Applications in the Social Scienes #163. Thousand Oaks: Sage Publications. Chapters 1-2.

Murtagh, Fionn (2005). *Correspondence Analysis and Data Coding with Java and R.* Chapter 1: Introduction. Pp. 1-28

Rouanet, Henry & Werner Ackermann and Brigitte Le Roux
"The Geometric Analysis of Questionnaires: The Lesson of Bourdieu's La Distinction." Pp. 5-18 in *Bulletin de Méthodologie Sociologique,* January 2000, N.65. 13 pages.

Lecture 2, Day 1.
    2. (Simple) Correspondence Analysis. Basic Concepts and Properties. Interpretation of Results.

Greenacre, Michael (1994). "Correspondence Analysis and its Interpretation." ." In Greenacre, Michael J. & Blasius, Jörg, eds., *Correspondence Analysis in the Social Sciences: Recent Developments and Applications.* San Diego: Academic Press. Pp. 3-22.

Greenacre, Michael J. (2007). *Correspondence Analysis in Practice.* London: Chapman & Hall. Chapter 9: Two-dimensional Displays, pp. 65-72

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple Correspondence Analysis.* Series: Quantitative Applications in the Social Scienes #163. Thousand Oaks: Sage Publications. Chapters 1-2.

Lecture 3, Day 1:
    3. From Simple to Multiple Correspondence Analysis.

Le Roux, Brigitte & Henry Rouanet (2004). *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis.* Dordrecht: Kluwer Academic Publishers.

Chapter 5: Multiple Correspondence Analysis. Pp. 214-241.
Chapter 6: Structured Data Analysis. Pp. 251-256, 265-268.
Chapter 9: Research Case Studies. Pp. 365-394.

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple Correspondence Analysis.* Series: Quantitative Applications in the Social Scienes #163. Thousand Oaks: Sage Publications. Chapter 3.

Henry Rouanet (2006). "The Geometric Analysis of Structured Individuals X Variables Tables." In Greenacre, Michael J. & Blasius, Jörg, eds. (2006). *Multiple Correspondence Analysis and Related Methods.* London: Chapman & Hall, pp.138-159.

Day 2:

Lecture 1:

    4. Multiple Correspondence Analysis. Specific Multiple Correspondence Analysis. Examination of The Cloud of Individuals.

Le Roux, Brigitte & Henry Rouanet:
"Interpreting Axes in Multiple Correspondence Analysis: Method of the Contributions of Points and Deviations." In Greenacre & Blasius (1998): *Visualization of Categorical Data.* San Diego: Academic Press, pp. 197-220.
23 pages.

Brigitte Le Roux and Henry Rouanet. *Geometric data analysis of individual differences*. pp. 1-36, 2003. http://epgy.stanford.edu/research/

Le Roux, Brigitte, Henry Rouanet, Mike Savage and Alan Warde (2008): "Class and Cultural Division in the UK." Under publication. http://www.cresc.ac.uk/publications/papers.html pp.1-29. 29 pages.

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple Correspondence Analysis.* Series: Quantitative Applications in the Social Scienes #163. Thousand Oaks: Sage Publications. Chapters 3, 4 and 6

Day 2, Lecture 2
5. **Stability Issues, Coding and Statistical Significance. Integration of Cluster Analysis and MCA.**

Le Roux, Brigitte & Henry Rouanet (2004). *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis.* Dordrecht: Kluwer Academic Publishers. Chapter 3: Euclidian Cloud. Pp. 106-116 & Chapter 9 : pp. 365-394

Murtagh, Fionn (2005). *Correspondence Analysis and Data Coding with Java and R.* Chapter 3: Input Data Coding. Pp. 71-110.

Romesburg, H. Charles (2004). *Cluster Analysis for Researchers.* North Carolina: LULU Press. Pp. 129-134.

Day 2, Lecture 3:

6. **Constructing Social Spaces.**

Bourdieu, Pierre (1984) *Distinction.* Cambr. Mass: Harvard University Press. Chapter 5: The Sense of Distinction. Chapter 6: Cultural Goodwill Appendix 1. Some Reflections on the Method. pp. 264-370 & pp.503-518.

Hjellbrekke, Johs. & al (2007). The Norwegian Field of Power anno 2004. In *European Societies*, Volume 9, Issue 2 May 2007 , pp. 245 – 273. 30 pp.

Le Roux, Brigitte and Henry Rouanet (2010). *Multiple Correspondence Analysis.* Series: Quantitative Applications in the Social Scienes #163. Thousand Oaks: Sage Publications. Chapter 6.

Lab sessions:

Brigitte Le Roux, Mikael Börjesson, Philippe Bonnet (2006):
Performing Multiple Correspondence Analysis (MCA) USING SPAD 1 (VERSION 6.5) Can be downloaded from
http://www.skeptron.uu.se/broady/sec/p-gda-0609-spadguide-mca.pdf (more material can be downloaded from the same site)

Data sets: TBA

SPSS 13.0. Categories. SPSS Inc.

Outline – lab sessions:

1.   Introduction to SPAD
     -   SPAD – import files
     -   setting up first analyses
     -   getting to know the software
     -   CA
     -   a first introduction to the output file
     -   interpretation of results
     -   Own analyses of other variables

2.   SPAD - MCA.
     -   MCA/MCAspe module in SPAD
     -   Detailed interpretation of results
     -   The graphic module in SPAD – how to manipulate the graphs
     -   First examination of the cloud of individuals
     -   Graphical tools of interpretation
     -   Use of supplementary variables

3.   SPAD – Cluster Analysis and MCA.
     -   More on SPAD – MCA/MCAspe module
     -   More detailed examination of the cloud of individuals
     -   Concentration, Indicator and confidence ellipses
     -   More on supplementary variable
     -   Integration of cluster analysis and MCA
     -   Interpretation of cluster analyses in the factorial planes