Oxford, July 2006

What is Multiple Correspondence Analysis?

[What is MCA?]

Brigitte Le Roux

MAP5 & CNRS

Mathematics & Informatics Department

Université René Descartes

lerb@math-info.univ-paris5.fr

http://math-info.univ-paris5.fr/~lerb/

1 Introduction

Language of questionnaire

Basic data set: Individuals×Questions table

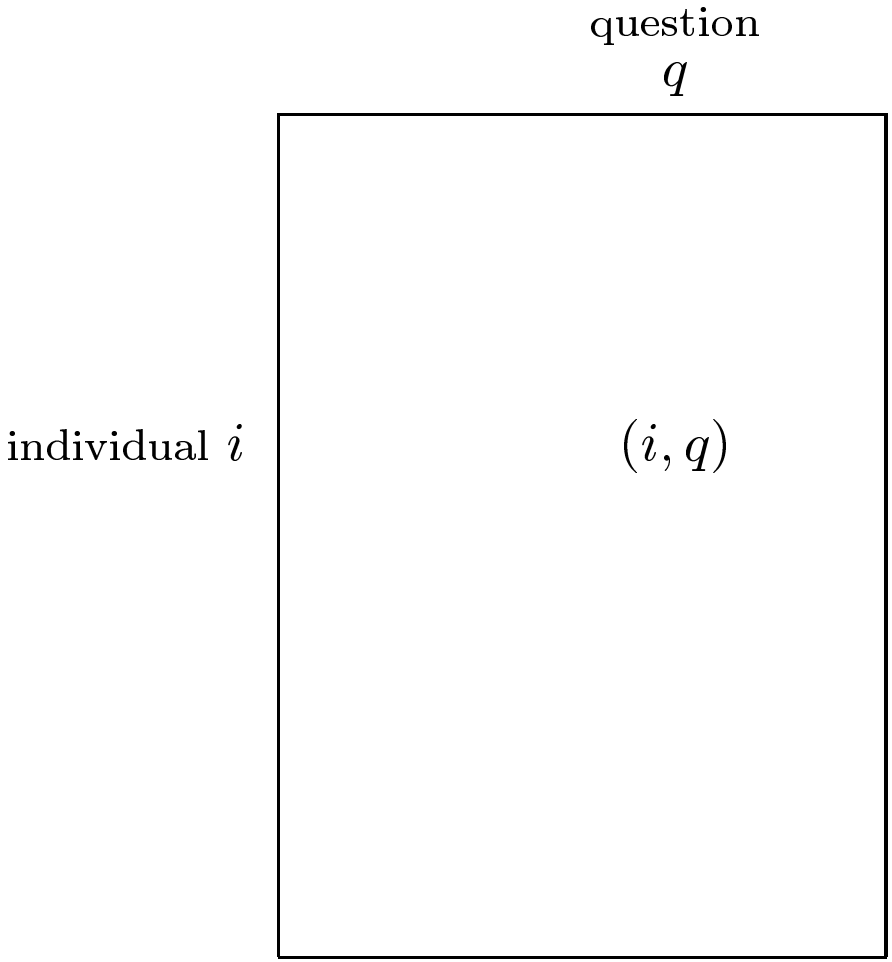
Questions are categorized variables, that is, variables with a finite number of categories, or modalities.

Questionnaire in "standard format": for each question, each individual chooses *one and only one* response modality.

I: set of n individuals

Q: set of questions

Basic data table analyzed by MCA:



1.1 Historical landmarks

Guttman (1941)

Burt (1950)

Benzécri (1972-1977)

Lebart (1975)

Bourdieu & Saint-Martin 1978 (*Le patronat*).

2 Principles of MCA

$$\text{MCA} \longrightarrow \begin{cases} \text{cloud of individuals} \\ \text{cloud of modalities} \end{cases}$$

Distance between two individuals i and i' for question q

i chooses modality k; i' chooses modality $k' \neq k$:

$$d_q^2(i, i') = \frac{1}{n_k/n} + \frac{1}{n_{k'}/n}$$

Distance between two modalities k and k'

$$d^2(k, k') = \frac{n_k + n_{k'} - 2n_{kk'}}{n_k\, n_{k'}/n}$$

n_k = number of individuals who have chosen k (resp. $n_{k'}$);
$n_{kk'}$ = number of individuals who have chosen both k and k'.

Principal axes, eigenvalues and contributions

Fundamental properties:

• the principal axes of the cloud of individuals are in a one-one correspondence with those of the cloud of modalities,

• the two clouds have the same eigenvalues.

Aids to interpretation: Contributions

Overall contribution of question q: $\frac{K_q - 1}{K - Q}$

(K_q: number of modalities of question q)

Contribution of point to axis: $\frac{p\, y^2}{\lambda}$

(y: coordinate of point on axis; p: relative weight)

Contributions add up by grouping $\rightarrow$ contribution of a question.

3 Steps of analysis

1. Choose active individuals, active questions (and coding);
2. Decide how many axes to be interpreted and interpret axes retained;
3. Investigate the cloud of individuals.

Case Study: The Norwegian field of Power[a]

Data from the survey on elites conducted by the *Norwegian Power & Democracy project* in Autumn 2000.

Research questions

• What are the different types of capital which separate between the different fractions in the Norwegian field of power?

• What fractions of this field are the most open with respect to social mobility?

• Are there particularly homogeneous fractions?

[a]See Hjellbrekke J., Le Roux B., Korsnes O., Lebaron F., Rosenlund L. & Rouanet H.: The Norwegian Field of Power Anno 2000. To appear in *European Societies*.

Data set

Active individuals: 1710 persons belonging to the following institutions:

Public & Private business	51+297
Public & Private cultural org./institutions	95+48
Political system (members of parliament + others)	138+62
Police and judicial system	78+60
Research and higher educational institutions	146
Central administration	197
Defence/Military	68
Church	107
Cooperatives	42
Media	116
Organizations	215

Choosing active questions and encoding modalities

- **Economic Capital.** 1 heading, 3 variables

(personal income, income on capital, registered property).

- **Educational Capital.** 2 headings

Personal: 3 variables

(own educational level, studies abroad, worked abroad);

Inherited: 2 variables

(father and partner educational level).

- **Social Capital.** 2 headings

Inherited (5 binary coded variables);

Personal (8 binary variables).

- **Professional experience.**: 10 binary coded variables

Rare modalities, non-responses, "junk" modalities

Rare modalities (say, of frequencies less that 5%) need to be pooled with others whenever feasible, or alternatively be put as "passive" ones (Specific MCA).

Technique of Supplementary Elements:

Supplementary variables

Supplementary individuals

Basic Results

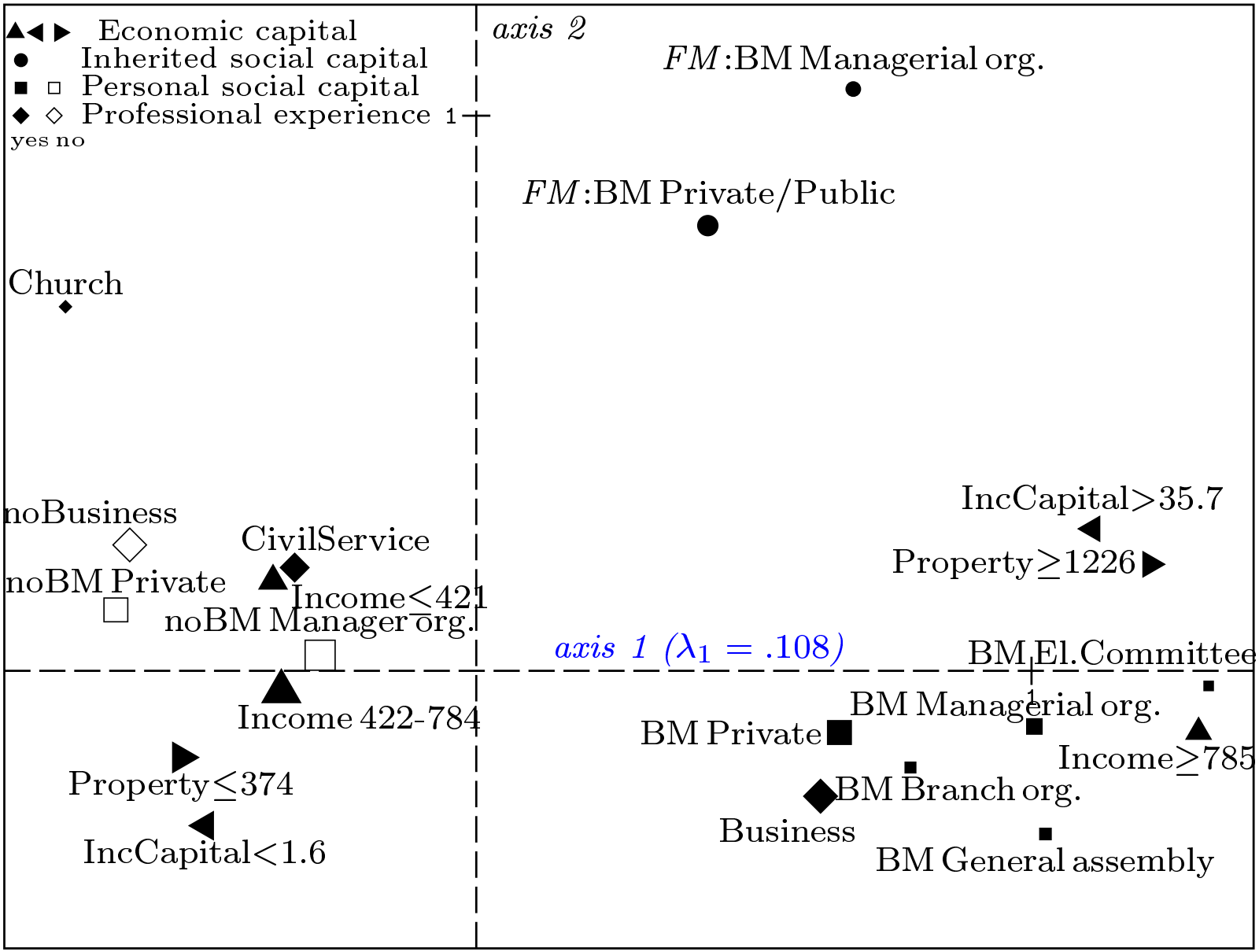
(i) the variances of axes (eigenvalues);
(ii) the principal coordinates of modalities and of individuals;
(iii) the contributions of categories to axes;
(iv) the geometric representation of the two clouds (modalities and individuals).

Interpreting axes

To give an assessment of the importance of axes, better than the raw rates, we calculate *modified rates* using the Benzécri's formula.

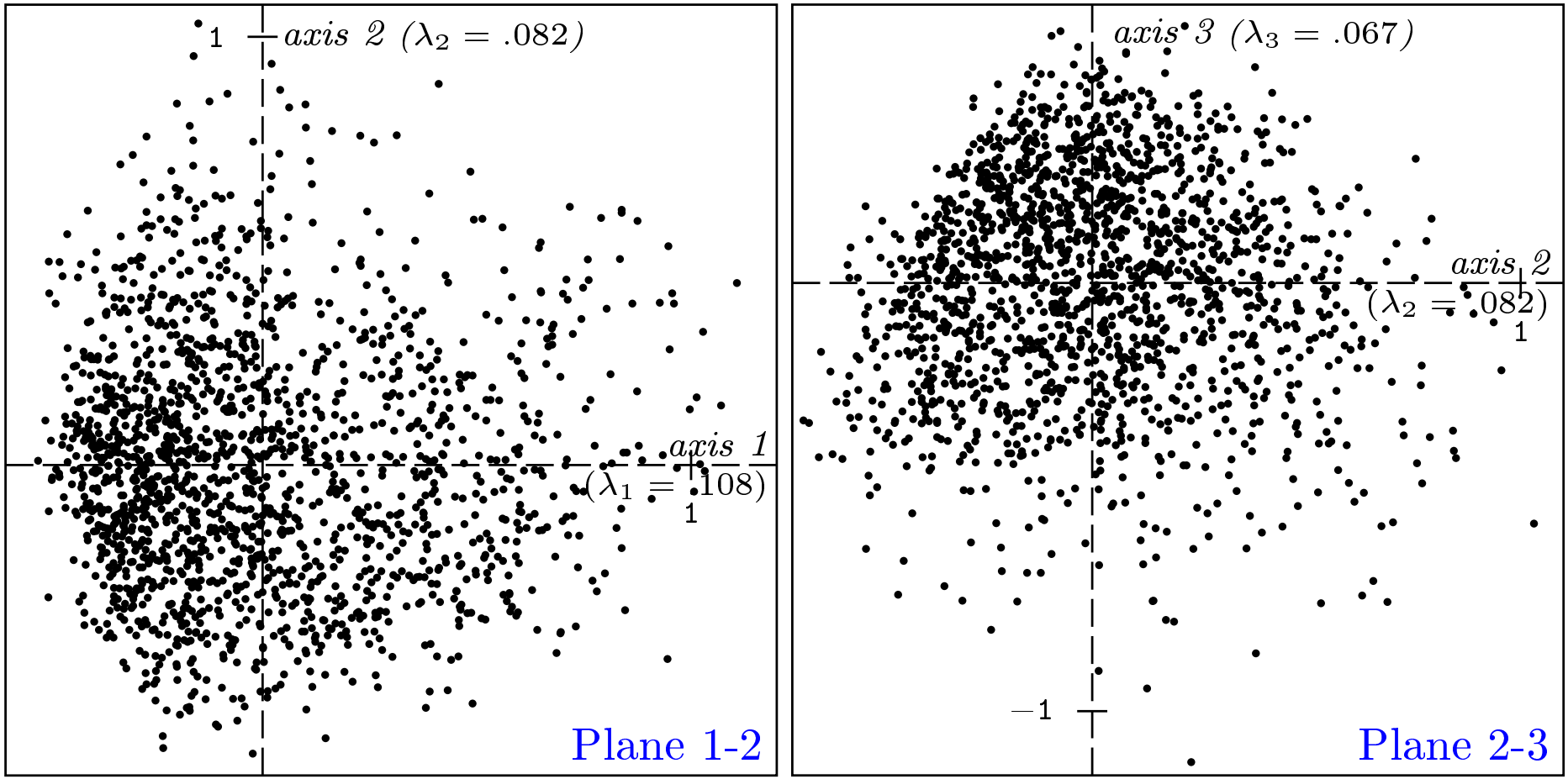
The interpretation of each axis is conducted in the cloud of modalities and based on the modalities whose contributions to axis exceed a specified threshold (average contribution).

Plane 1-2: 20 modalities contributing to axis 1 (4 headings), marker sizes proportional to frequencies (*FM*= Father/Mother, BM=Board Member).



Exploring the cloud of individuals

Cloud of 1710 individuals in plane 1-2 and in plane 2-3.

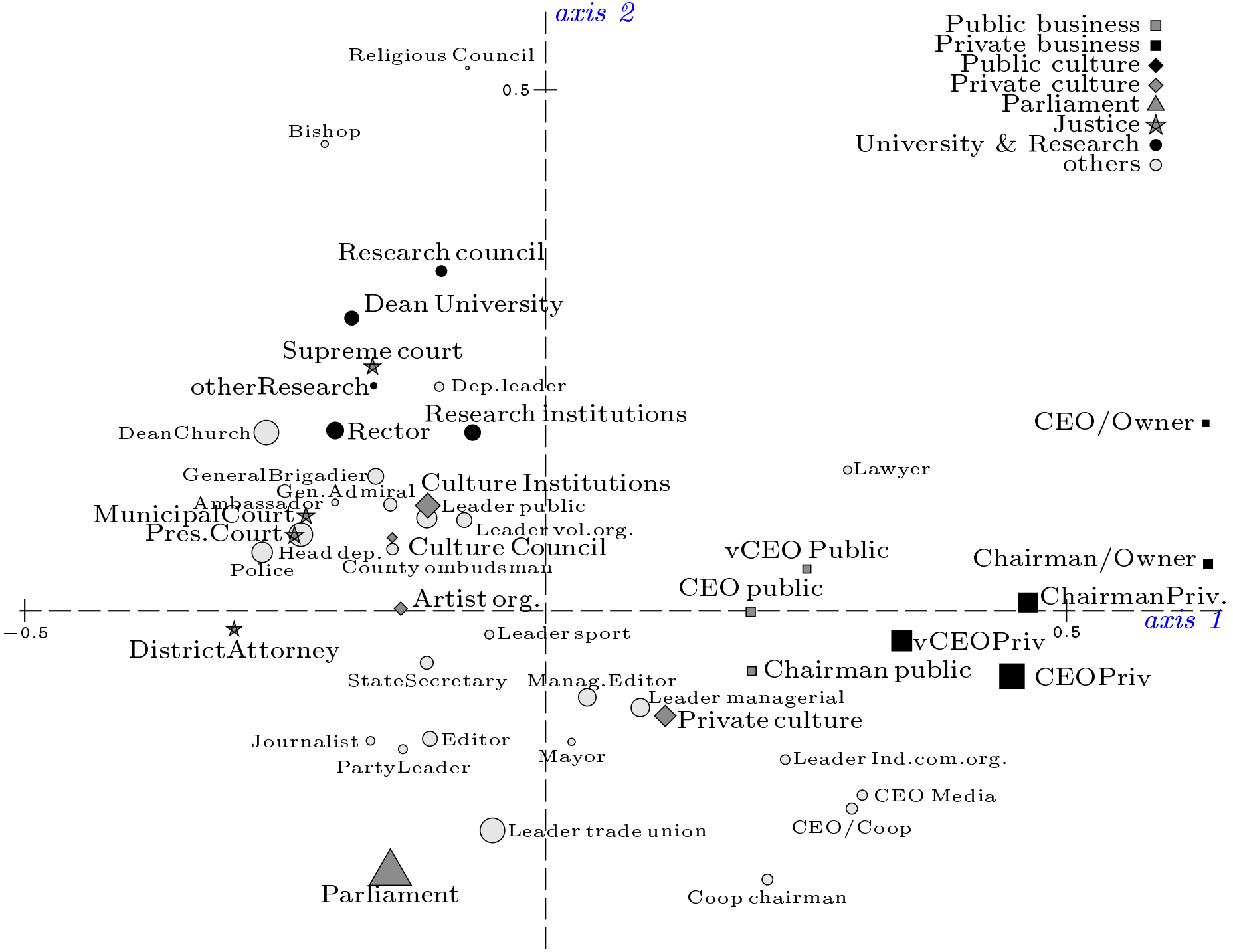


Subclouds and modality mean points

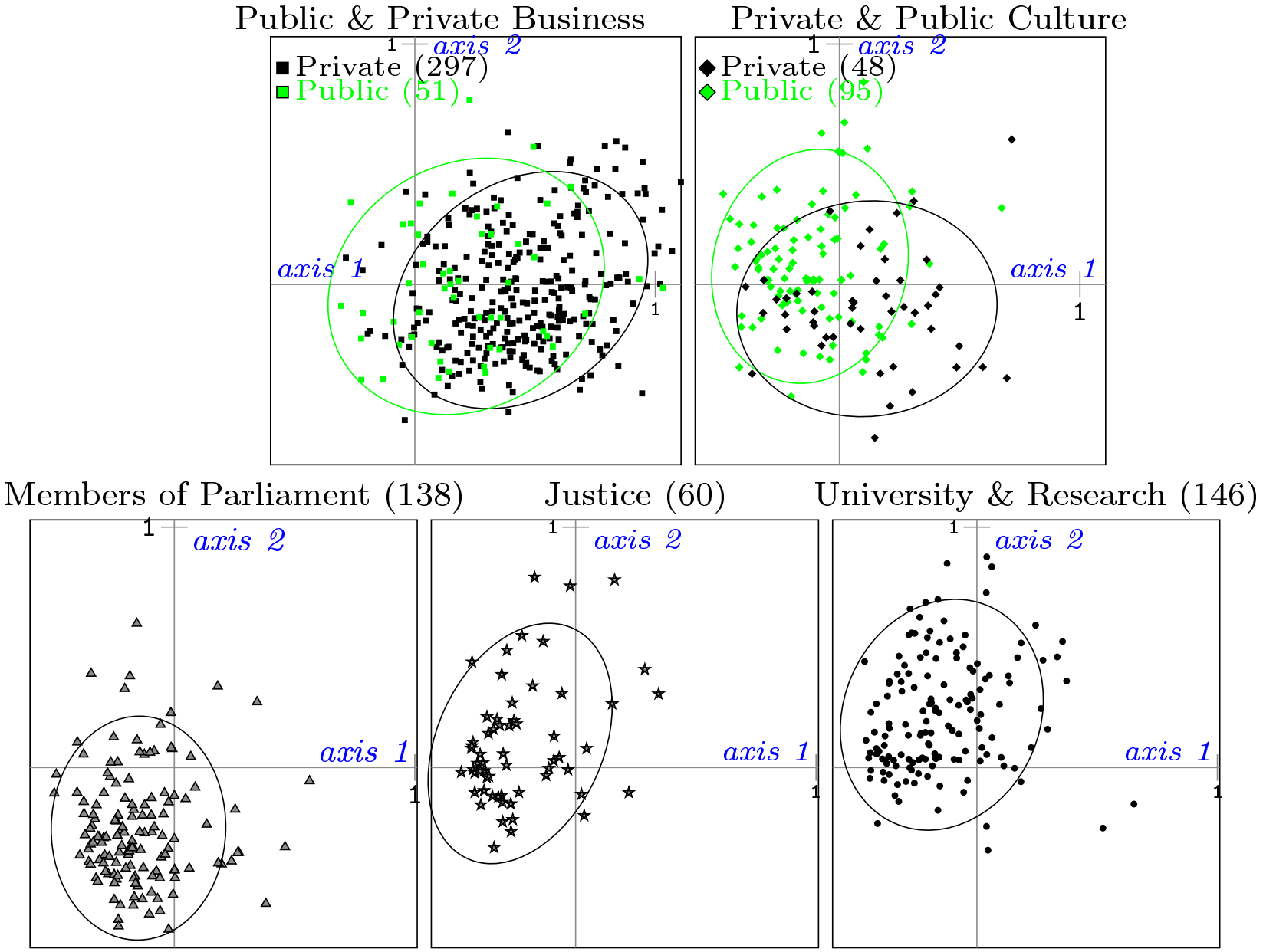
Subcloud of the individuals having chosen one modality $\rightarrow$ its mean point is called a *modality mean-point*.

Coordinate of the modality mean-point $= \sqrt{\lambda}\, y$

(y is the coordinate of the modality in the space of modalities)

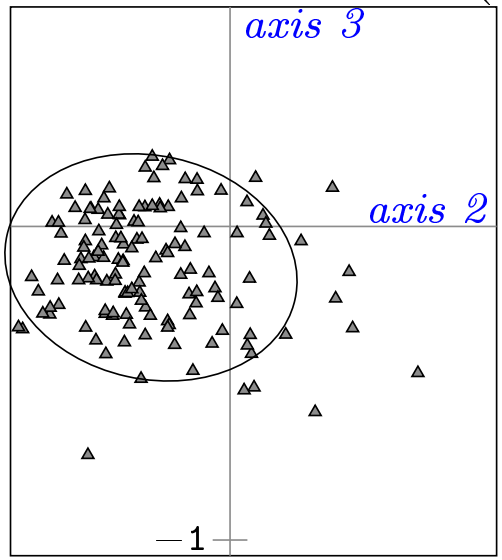

axis 2
Public business
Private business
Public culture
Private culture
Parliament
Justice
University & Research
others
Religious Council
0.5
Bishop
Research council
Dean University
Supreme court
otherResearch
Dep.leader
Research institutions
DeanChurch
Rector
CEO/Owner
Lawyer
GeneralBrigadier
Gen.Admiral
Culture Institutions
Ambassador
Leader public
MunicipalCourt
Pres.Court
Leader vol.org.
Culture Council
Head dep.
County ombudsman
Police
vCEO Public
Chairman/Owner
CEO public
ChairmanPriv.
Artist org.
axis 1
−0.5
0.5
Leader sport
DistrictAttorney
vCEOPriv
Chairman public
CEOPriv
StateSecretary
Manag.Editor
Leader managerial
Private culture
Journalist
Editor
Mayor
Leader Ind.com.org.
PartyLeader
CEO Media
CEO/Coop
Leader trade union
Parliament
Coop chairman

Concentration ellipses

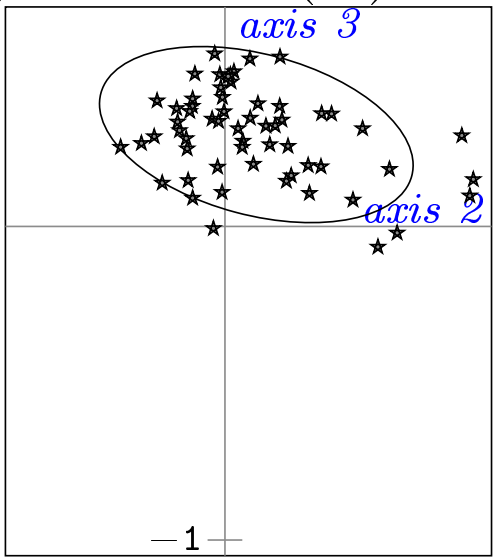


Concentration ellipses of 3 subgroups of interest in plane 2-3

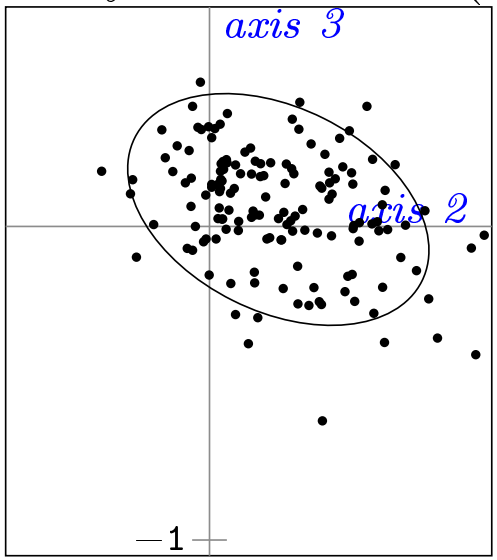
Members of Parliament (138)



Justice (60)



University and Research (146)



4 Final message about MCA

MCA is a method for the *Geometric Analysis* of questionnaires

Investigating the *cloud of individuals*: individuals carry all the information!

Theoretical references about MCA

— Benzécri J-P. (1977). Sur l'analyse des tableaux binaires associés à une correspondance multiple [On the analysis of binary tables associated with a multiple correspondence], *Les Cahiers de l'Analyse des Données*, 2, 55-71 (from a mimeographed note of 1972).

— Escofier B. & Pagès J. (1988), *Analyses factorielles simples et multiples* [Simple and Multiple factor Analyses] (chapter 3 on mca). Paris: Dunod.

— Greenacre M. (1984). *Theory and Applications of Correspondence Analysis* (chapter 5 on mca). London: Academic Press.

— Lebart L., Morineau A. & Warwick K.M. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices* (chapter 4 on mca). New York: Wiley.

— Le Roux B., & Rouanet H. (2004). *Geometric Data Analysis; From Correspondence Analysis to Structured Analysis*. Dordrecht: Kluwer (the chapter 5 contains a detailed presentation of mca and an extensive illustration: "the Culture Example")

Recent substantive studies

— Bonnet P., Le Roux B. & Lemaine G. (1996). Analyse géométrique des données: une enquête sur le racisme [Geometric Data Analysis: a survey on racism], *Mathématiques et Sciences Humaines*, 136, 5-24.

— Bourdieu P. (1999). Une révolution conservatrice dans l'édition [A conservative revolution in publishing], *Actes de la Recherche en Sciences Sociales*, Vol. 126-127, 3-28.

— Chiche J., Le Roux B., Perrineau P. & Rouanet H. (2000). L'espace politique des électeurs français à la fin des années 1990 [The political space of French electors in the late 1990s]. *Revue française de sciences politiques,* 50, 463-487.

— Hjellbrekke j., Le Roux B., Korsnes O., Lebaron F., Rosenlund L. & Rouanet H. (to appear). The Norwegian field of Power Anno 2000. *European Societies.*

— Le Roux B. & Rouanet H. (2003). Geometric Analysis of Individual Differences in Mathematical Performance for epgy Students in the Third Grade. `http://www-epgy.stanford.edu/research/`.

— See chapter "Case Studies" in Le Roux & Rouanet (2004) (op. cit.)

About software

These results have been obtained using ADDAD, ellipse and EyeLID software freely avalaible from my website `math-info.univ-paris5.fr/~lerb/` under the "Logiciels" heading.

All results presented here will be obtainable from the September 2006 version of SPAD software distributed by SPAD firm (www.spadsoft.com).