groStat 2012

A new proposal, Multiway Discriminant Analysis: STATIS-LDA

Robert Sabatier, Myrtille Vivien & Christelle Reynès
EA 2415

Université Montpellier I

Regional wines from J. Pagès

Pagès J, Analyse Factorielle Multiple, Agrocampus document, Rennes.

The Data:

n = 21 wines, *I* = 3 origins : Saumur (11), Chinon (4), Bourgueil (6)

29 variables (average rates from 36 judges) organized into *K* = 5 tables

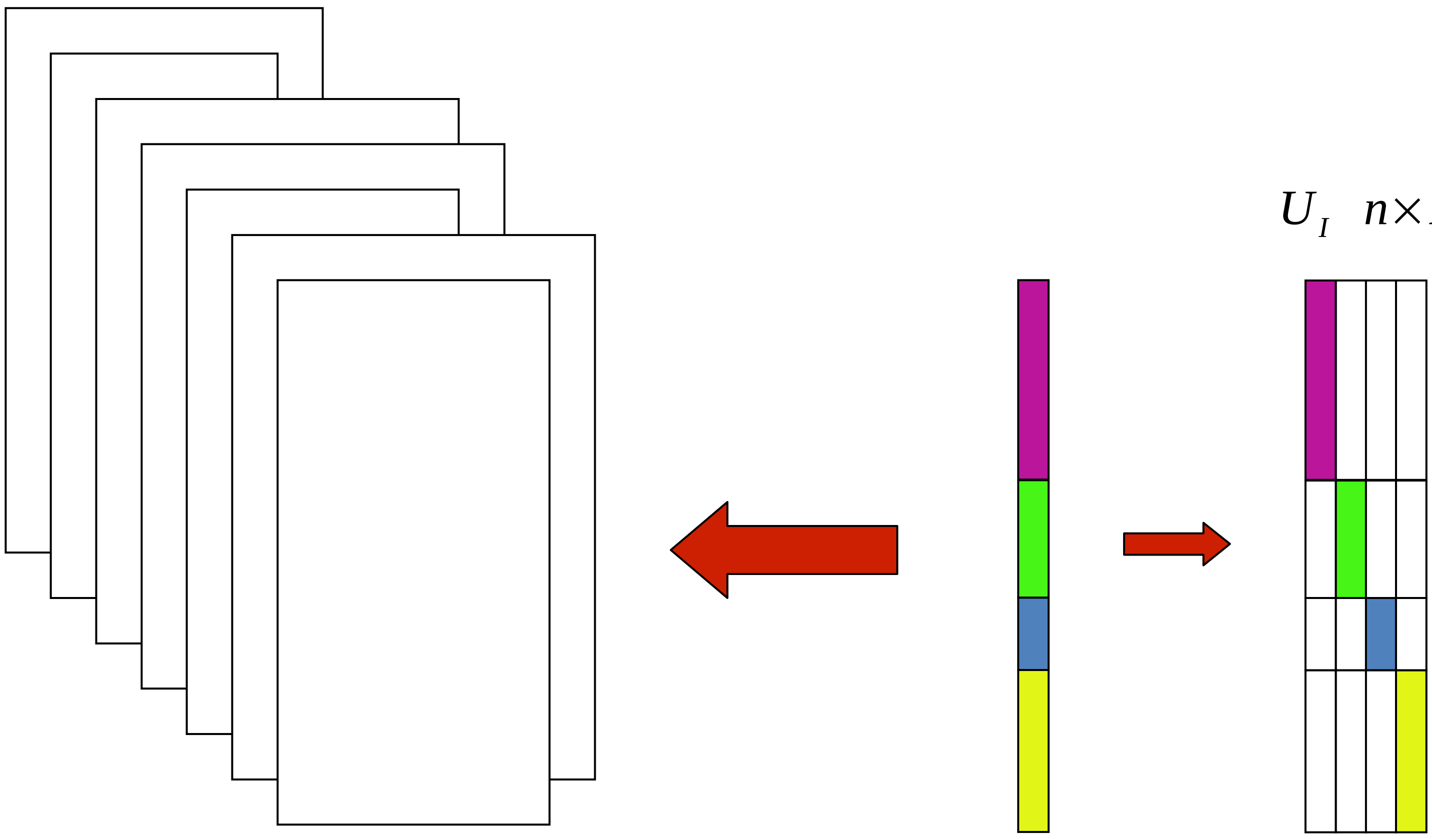
1. Olfaction before stirring (5 variables)
2. Vision (3 variables)
3. Olfaction after stirring (10 variables)
4. Taste (9 variables)
5. Overall judgement (2 variables)

Objectives:

Showing the main dimensions of **sensorial variability of wines**.
Relating those dimensions to **origin**

Data Structure

1 multiway table *: K* tables X_k $n \times p_k$



Discrimination for multi-way tables

- **Objective** : predict the class of new observations

  → use of the **within- and between-variances**

  → impossible use of PLS-like ou N-PLS, ACG or Generalized Procruste

- the **same classes** for all tables

- **This is not a cube**

  → methods like PARAFAC/CANDECOM, TUCKER… are inappropriate

- We want to **« keep » the multiway structure** and not to joint tables into one matrix

STATIS-LDA →

- Keeps between/within decomposition
- Results independent on units
- Keeps the data structure
- Works on high-dimensional data
- Allows to quantify individual table importance
- Allows validation processes

Reminder about STATIS (ACT)

- Notations

$$\begin{cases} (X_k, Q_k, D) & \text{PCA triplet of the k-th table D-centred} \\ Q_k & \text{Square metric to compute distances between observations} \\ D & \text{Observation Weighting} \\ W_k D = X_k Q_k X_k^t D & \text{Scalar product between observations} \\ W_c D = \sum_{k=1}^{K} \nu_k W_k D & \text{Compromise operator found by STATIS} \end{cases}$$

- Definition : $(X_k, Q_k, D)_{k=1,\dots,K}$ STATIS

$$\text{Finding } W_c D \text{ verifying } \underset{\nu^t \nu = 1}{Max}\left\{ \left\| W_c D \right\|^2 = tr\left(W_c D W_c D\right) \right\}$$

- Solution :

$$C\nu = \lambda \nu \text{ with } C = \left[c_{k,k'} = tr\left(W_k D W_{k'} D\right) \right] \Rightarrow \begin{cases} \text{interstructure} \\ \text{compromise} \\ \text{intrastructure} \end{cases}$$

Reminder about LDA of X_k with regards to U_I

- Notations

$$\begin{cases} V_k = X_k^t D X_k & \text{total variance matrix} \\ D_I = U_I^t D U_I & \text{class frequency matrix} \\ G_k = D_I^{-1} U_I^t D X_k & I \times p_k \text{ matrix of class baricenters} \\ V_{B_k} = G_k^t D_I G_k & \text{between-class variance} \\ V_{W_k} = V_k - V_{B_k} & \text{within-class variance} \\ P_{U_I} & D-\text{orthogonal projector onto } \mathrm{Im}(U_I) \\ P_k & D-\text{orthogonal projector onto } \mathrm{Im}(X_k) \end{cases}$$

- Definition : **LDA (according to Fisher)**

$$\text{Finding } c = X_k a_{LDA} \text{ verifying } \underset{a_{LDA}}{Max} \left\{ \frac{a_{LDA}^t V_{B_k} a_{LDA}}{a_{LDA}^t V_{W_k} a_{LDA}} \right\}$$

- Equivalent solutions :

$$V_{W_k}^{-1} V_{B_k} a_{LDA} = \lambda a_{LDA}$$

axes induced from $\left(G_k, V_k^{-1}, D_I\right)$ PCA ones, with $a_{LDA} = \frac{1}{\lambda_{ACP}} V_k^{-1} a_{ACP}$

c (observations coordinates) obtained by diagonalizing $P_k P_{U_I}$

STATIS-LDA definition

- Definition: it is STATIS on $\left(G_k, V_k^{-1}, D_I\right)_{k=1,\ldots,K}$

- A few properties:

If $K=1$, STATIS-LDA is LDA

$$C_{K\times K} = \left[c_{k,k'} = tr\left(W_k D W_{k'} D\right) = tr\left(V_{B_{kk'}} V_{k'}^{-1} V_{B_{k'k}} V_k^{-1}\right) = tr\left(P_{U_I} P_k P_{U_I} P_{k'}\right)\right]$$

$W_c D$ is solution of $\underset{\nu^t\nu=1}{Min}\left\|\sum_{k=1}^{K} \nu_k P_k P_{U_I}\right\|$

$W_c D = \sum_{k=1}^{K} \nu_k G_k V_k^{-1} G_k^t D_I$, linear combination of the K LDA $W_k D$

$W_c D$ the WD operator of the triplet $\left(\tilde{G}, \tilde{V}^{-1}, D_I\right)$

with $\tilde{X} = \left[X_1, X_2, \ldots, X_K\right]$, $\tilde{G} = D_I^{-1} U_I D \tilde{X}$ and $\tilde{V}^{-1} = diag\left(\nu_k V_k^{-1}\right)$

It is not LDA but a **modified LDA**

STATIS-LDA implementation

- Inside the algorithm: two diagonalizations : $C \ (K \times K)$ and $W_c D \ (I \times I)$

- Basic outputs:

$$\begin{cases} C & \text{matrix of the scalar products between the } K \text{ operators} \\ \{v_\alpha\}_{\alpha=1,\dots,K} & \text{interstructure (} K \text{ operator representation)} \\ \{c_\alpha^{cdg}\}_{\alpha=1,\dots,A} & \text{coordinates of the baricenters representation} \end{cases}$$

- Other outputs:

$$\begin{cases} \{a_\alpha\}_{\alpha=1,\dots,A} & \text{STATIS-LDA axes: } a_\alpha = \dfrac{1}{\mu_\alpha(W_c D)} \tilde{V}^{-1} c_\alpha^{cdg} \\ \{c_\alpha\}_{\alpha=1,\dots,A} & n \text{ observations coordinates : } c_\alpha = \tilde{X} a_\alpha \\ \{c_{k_\alpha}\}_{\alpha=1,\dots,A} & \text{coordinates of observations with regards to } X_k : c_{k_\alpha} = X_k a_\alpha \end{cases}$$

- Graphical representations of observations and variables
- Global and table-wise good classification rates
- Cross-validation and test sets

Regional wines from J. Pagès

Pagès J, Analyse Factorielle Multiple, Agrocampus document, Rennes.

The Data:

n = 21 wines, I = 3 origins : Saumur (11), Chinon (4), Bourgueil (6)

29 variables (average rates from 36 judges) organized into K = 5 tables

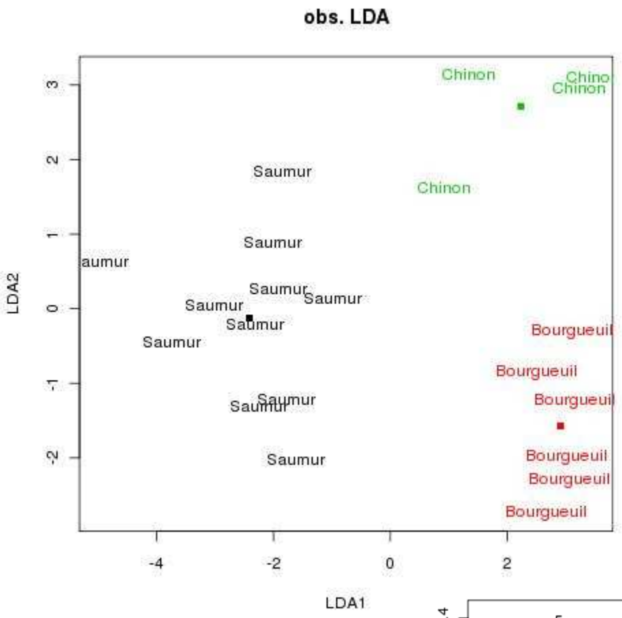
1. Olfaction before stirring (5 variables)
2. Vision (3 variables)
3. Olfaction after stirring (10 variables)
4. Taste (9 variables)
5. Overall judgement (2 variables)

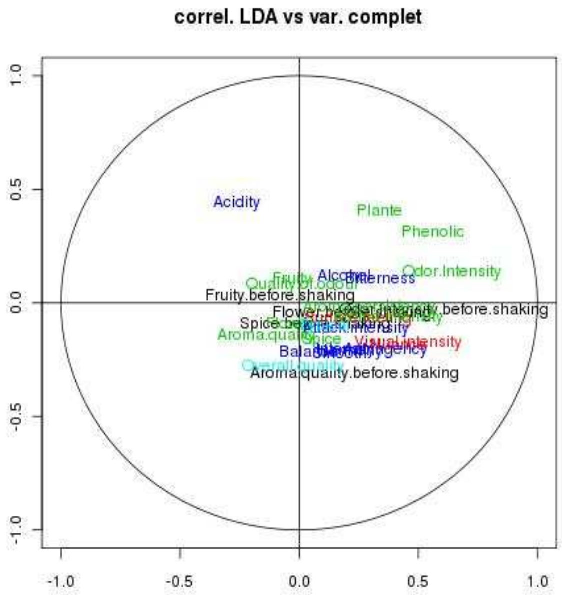
Objectives:

Showing the main dimensions of **sensorial variability of wines**.
Relating those dimensions to **origin**

A few preliminary analyses

- LDA (global)

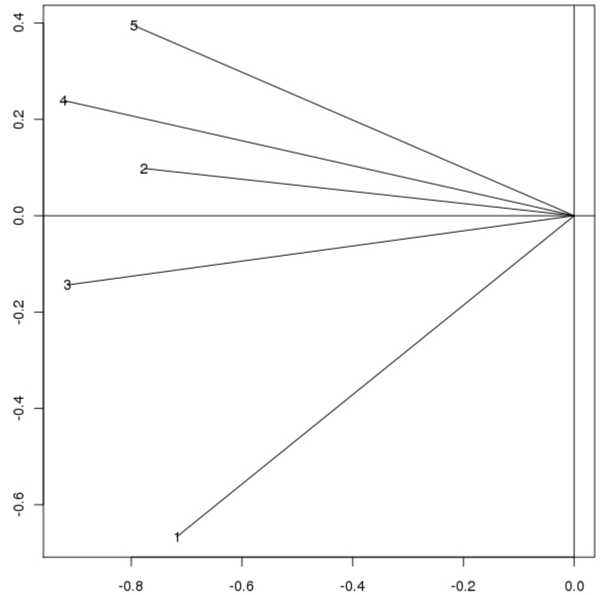




Discriminative ability of the first two axes: 0.882 and 0.712

- STATIS (usual)

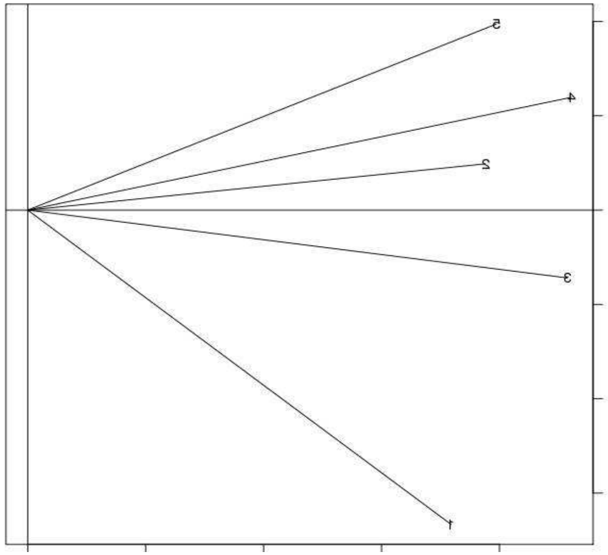
$$\left(X_k, Id, \frac{1}{21} Id\right)_{k=1,\dots,5}$$



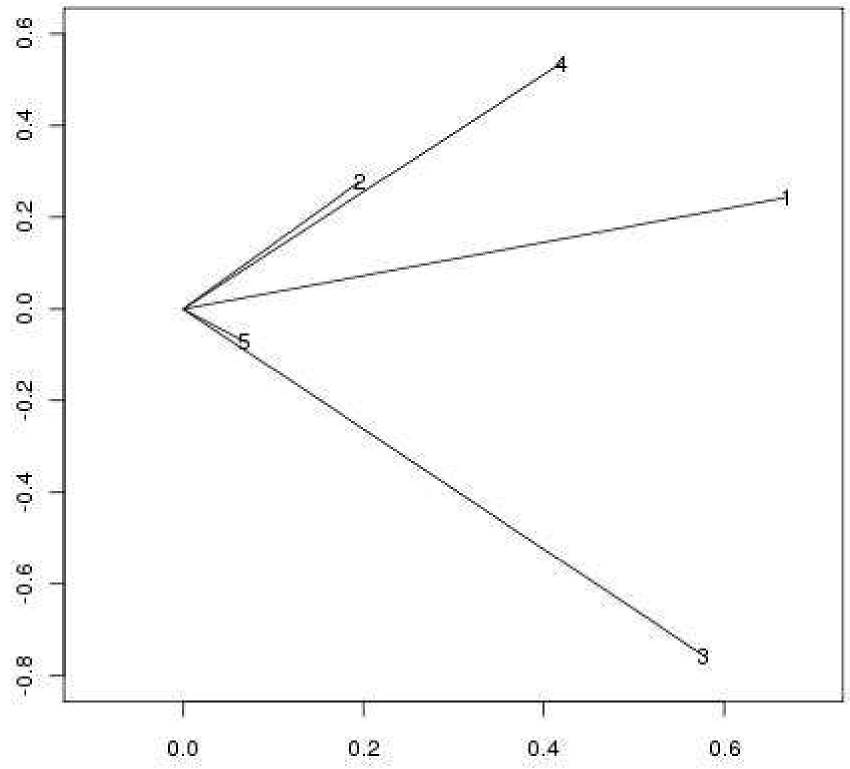
Global STATIS-LDA results

Interstructure

STATIS



STATIS-LDA



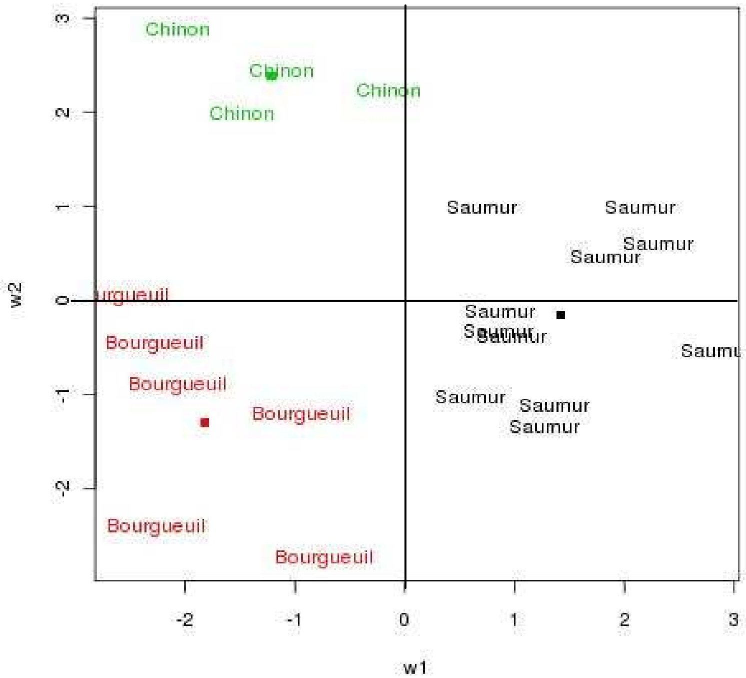
→ more relevent use of the different tables

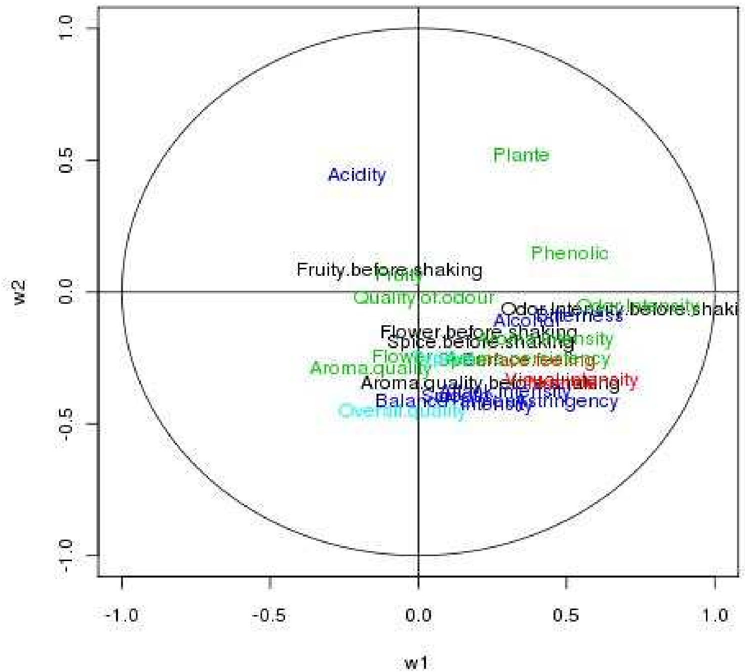
Interstructure coefficients (ν_k) :

1.40, 0.41, **1.21**, 0.88, 0.14

Résultat global STATIS-LDA

Représentation of compromise observations and variables





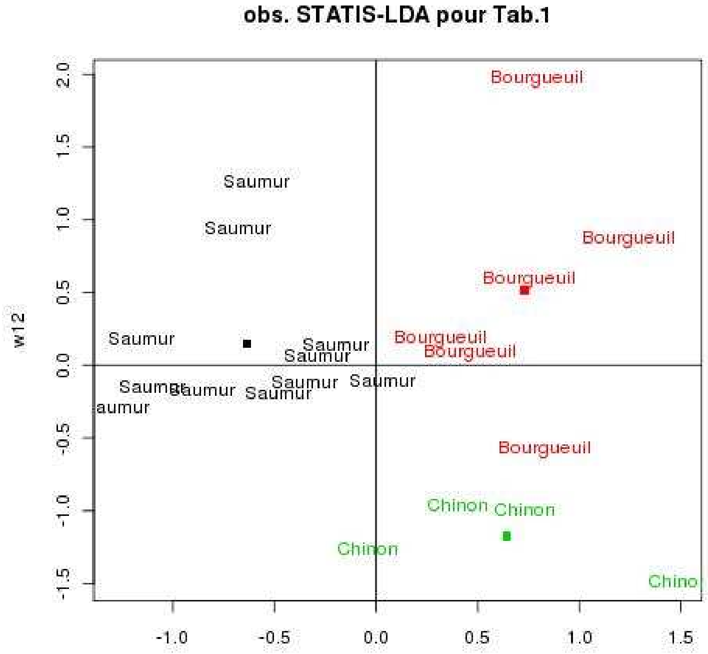
Discriminant ability: 0.821and 0.713

LDA : 0.882 and 0.712

→ a little global loss with regard to joint-LDA

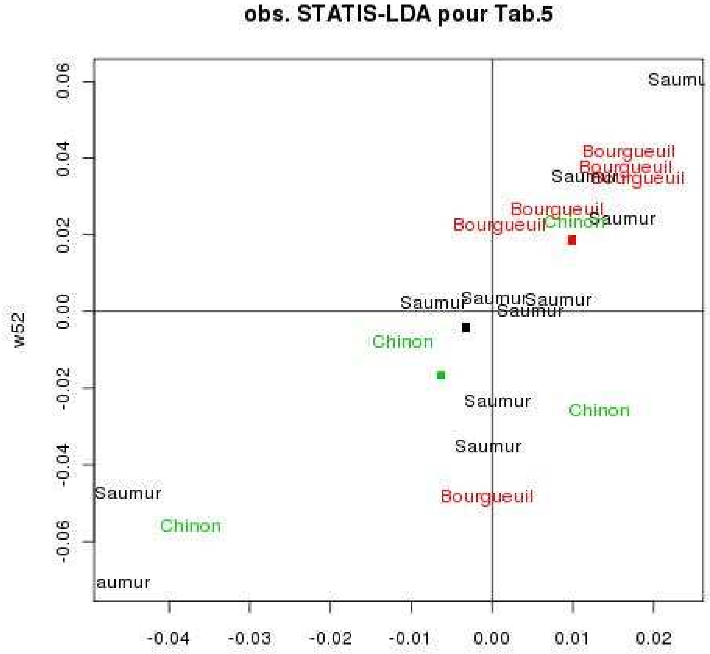
Table-wise STATIS-LDA results

Representation of intra structure observations



Discriminant ability: 0.722 and 0.534

LDA: 0.736 and 0.505



Discriminant ability : 0.107 and 0.126

LDA: 0.130 and 0.011

Insights in robustness and stability

Use of cross-validation

10-Fold cross-validation (run 100 times)

	LDA (global)	STATIS-LDA (partial)				
		X_1	X_2	X_3	X_4	X_5
Gcr	63.8%	**88.6**%	39.0%	46.7%	25.2%	34.8%

Addition of a 6-th « noisy » table

Addition of 3 independent variables of gaussian noise

		X_1	X_2	X_3	X_4	X_5	X_6
compromise	with noise	1.409	0.412	1.216	0.884	0.144	**0.023**
coef	original	1.401	0.409	1.209	0.878	0.143	

Wine exemple conclusion

- Exhibits the « discriminant » structure of the multi-way table, very different from the common structure.

- Allows to identify important tables for discrimination

- Obtains results close to LDA

Second example: Digits recognition on envelopes

Breukele, M, Duin, RPW, Tax, DMJ, Hartog, JE (1998) Handwritten digit recognition by combined classifiers, Kybernetika, 4, 381-386..

The data:

2000 digits written on envelopes (200 samples for each digit between 0 and 9)
- 1500 observations for the train set
- 500 observations for the test set

10 classes

For each observations, **six kinds of variables** are measures => **6 tables** :

1. fou: 76 Fourier coefficients measured on digits (2D descriptors),
2. fac: 216 correlation profiles,
3. kar: 64 Karhunen-Loève coefficients,
4. pix: 240 average pixels for 2 x 3 windows (initially 30 x 48 = 1440 pixels),
5. zer: 47 Zernike moments,
6. mor : 6 morphological variables.

Observation sample

Objectifs :

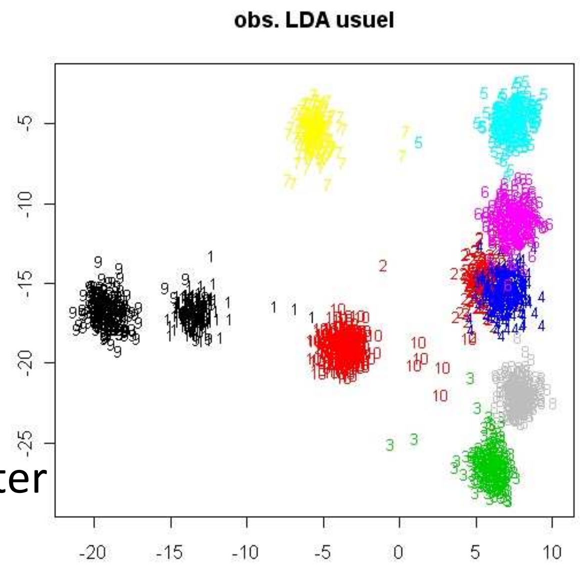
How the 6 sets of variables can be used to predict the digits?

Some preliminary analyses

- Global LDA (after collinear variables removal)

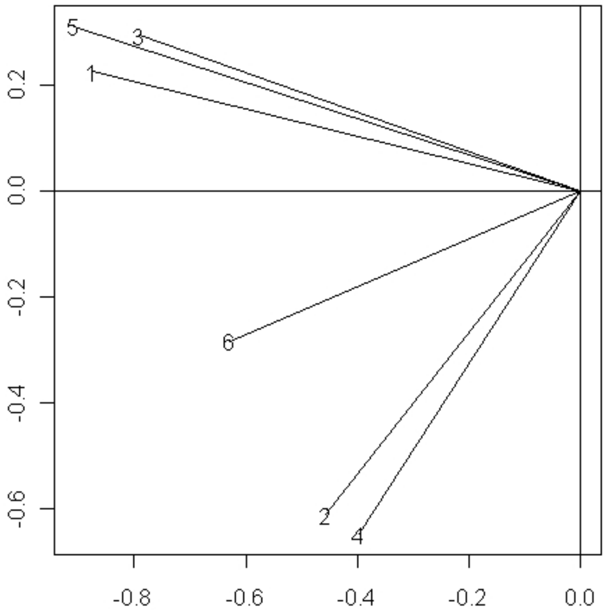
Learning set:
no bad classification out of 1500 !

Test set:
6 bad classifications out of 500



% good classifications on test set, **table by table**, after collinear variables removal:

Tab.	1	2	3	4	5	6
	0.972	0.810	0.948	0.722	0.954	0.822



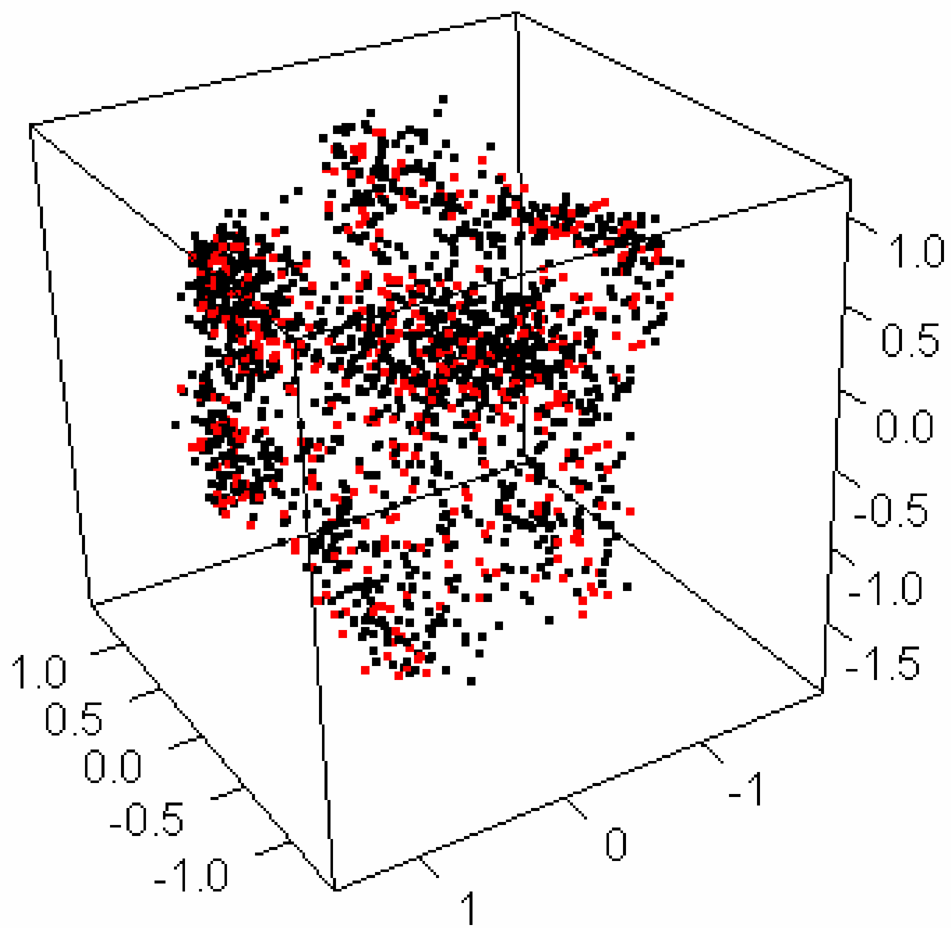
- STATIS (usual)

$$\left(X_k, Id, \frac{1}{1500} Id \right)_{k=1,\dots,6}$$

Some preliminary analyses

- STATIS (usual)

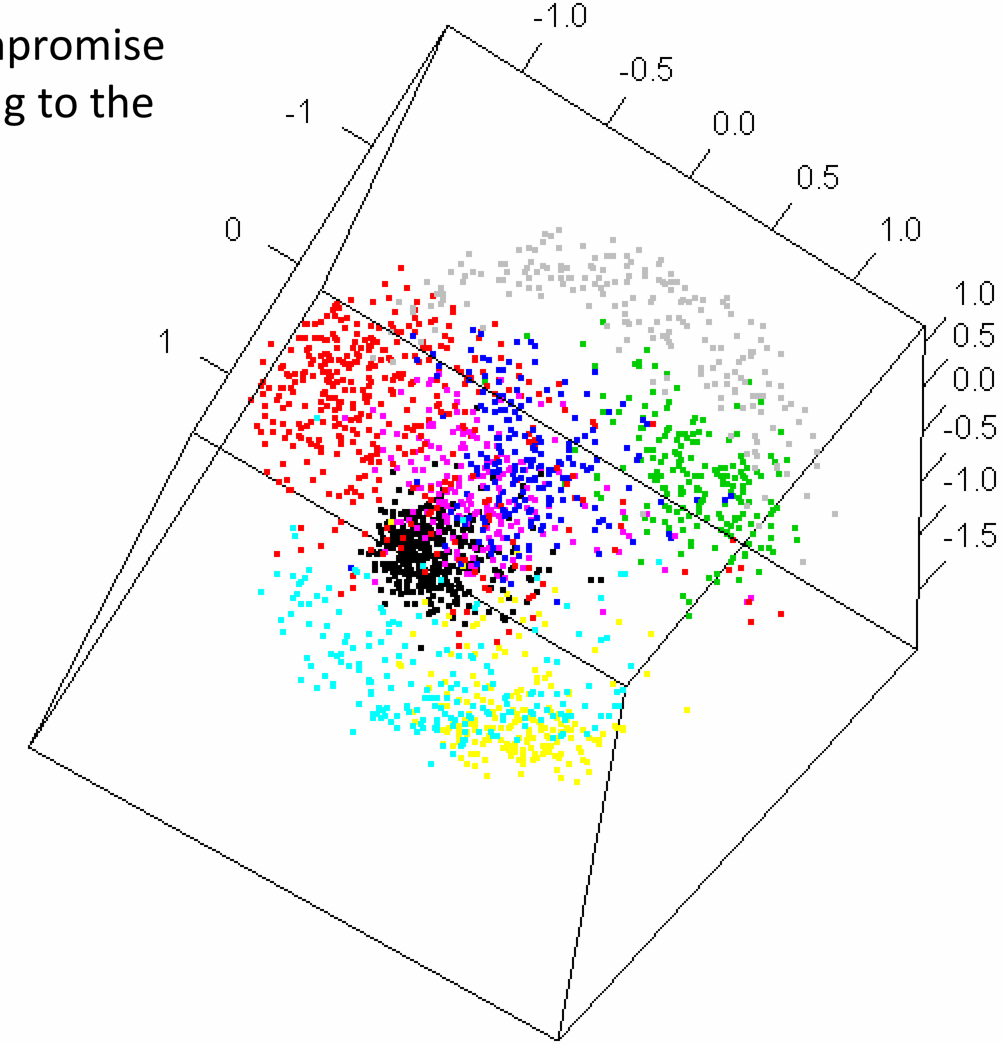
Representation of test observations in the compromise observation space



Some preliminary analyses

- STATIS (usual)

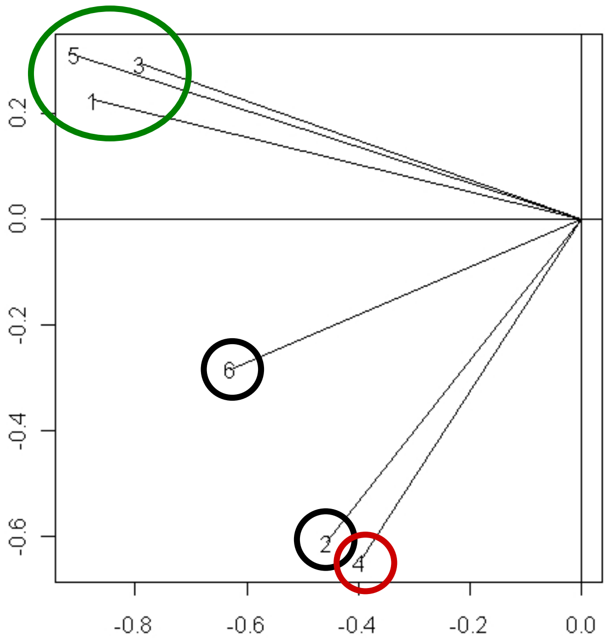
Representation of compromise observations (according to the class)



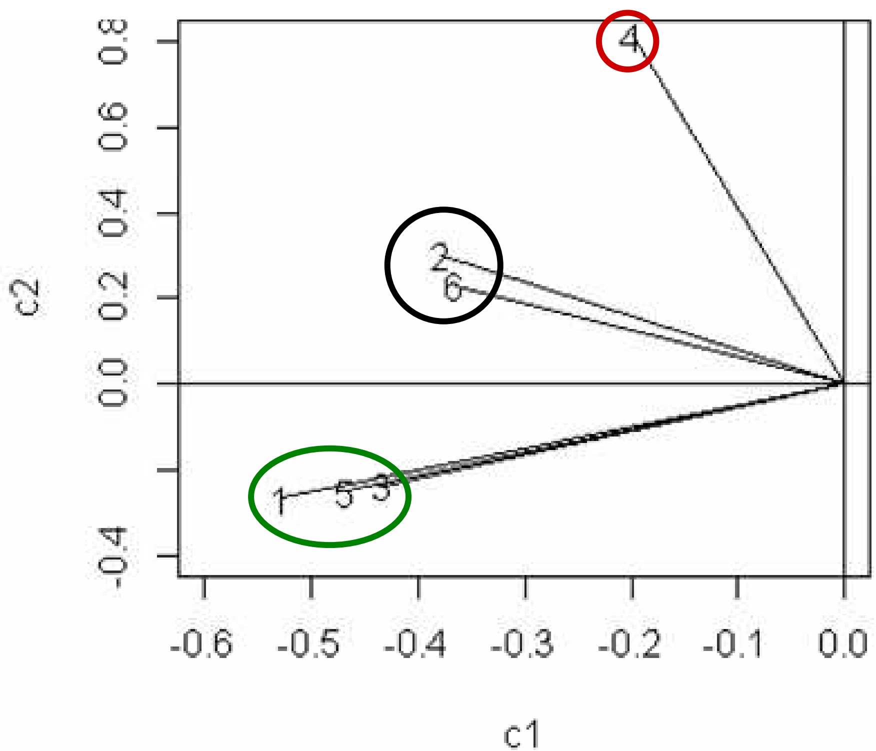
STATIS-LDA results

Tab.	1	2	3	4	5	6
gcrLDA	0.978	0.817	0.954	0.695	0.958	0.803
coefinterSTATIS-LDA	1.330	0.951	1.086	0.500	1.174	0.915

STATIS Interstructure

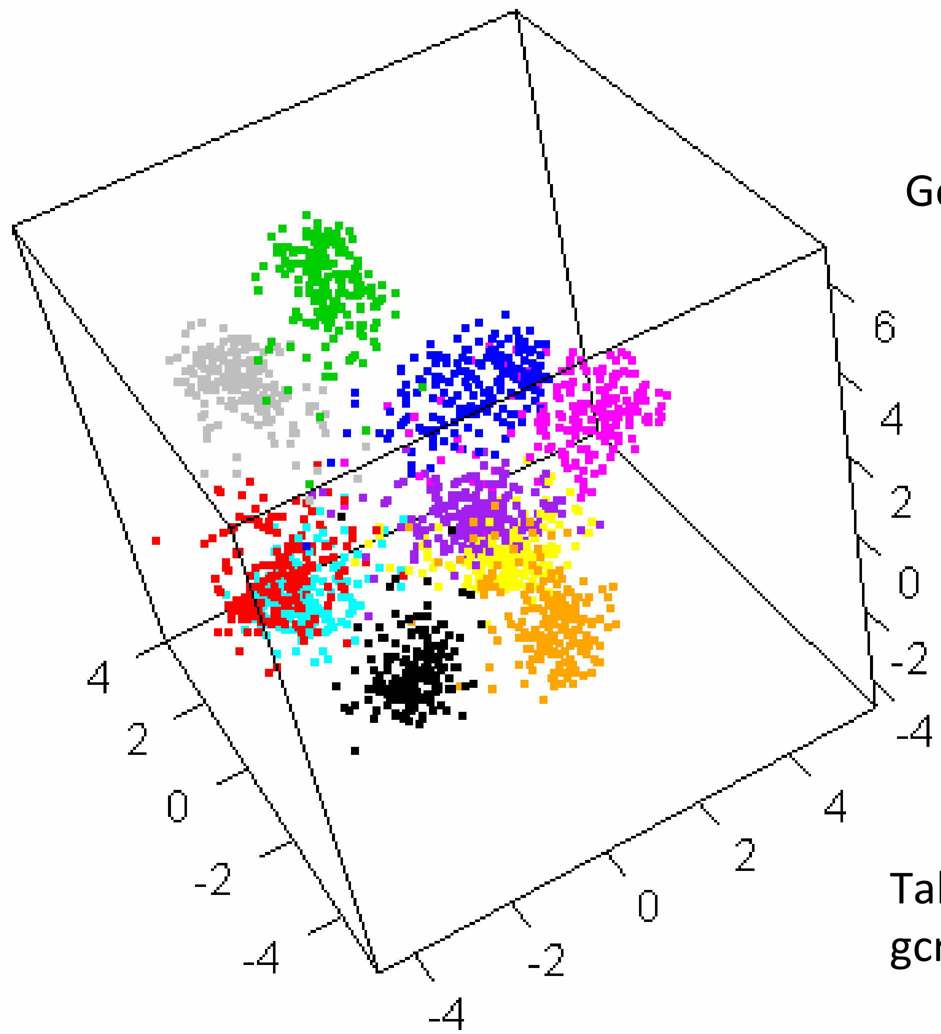


STATIS-LDA Interstructure



STATIS-LDA results

Compromise observations representation



Good classification rates:

Train: 98.6%
Test: 98%

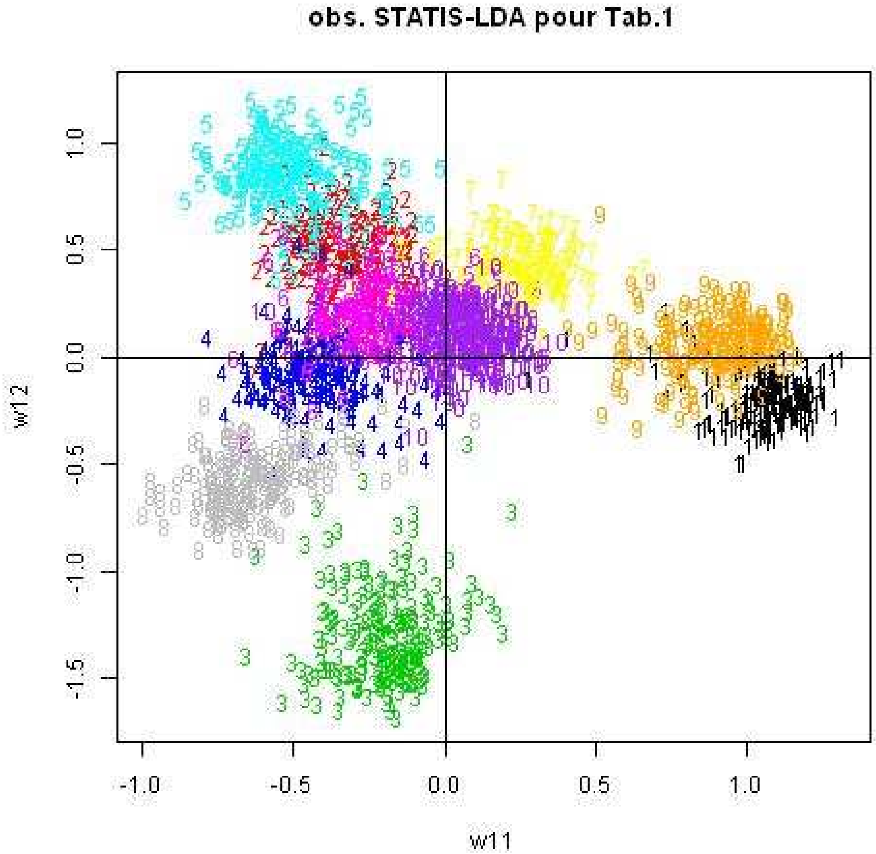
Partial results:

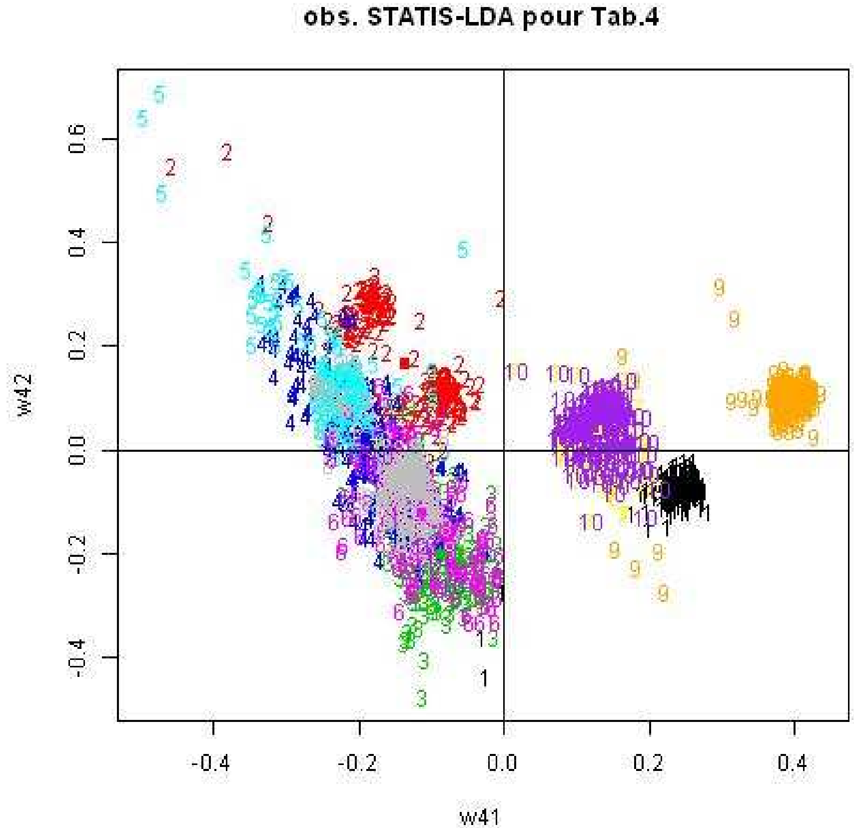
Tab.	1	2	3	4	5	6
gcr	0.978	0.817	0.954	0.695	0.958	0.803

STATIS-LDA results

Partial analysis :

Tab.	1	2	3	4	5	6
Indiviudal LDA gcr	0.978	0.817	0.954	0.695	0.958	0.803
Partial STATIS-LDA gcr	0.987	0.830	0.951	0.707	0.969	0.826





STATIS-LDA results

STATIS-LDA = precious tool for interpretation of classification in multiway table context

- interstructure: a good insight in tables relationship according to their role in classification
- Compromise coefficients: a quantification of table importance (see later)

BUT NOT ONLY…

STATIS-LDA results

Data modification: table 2 is kept and the other ones are randomly column-wise permutated

Partial-LDA results:

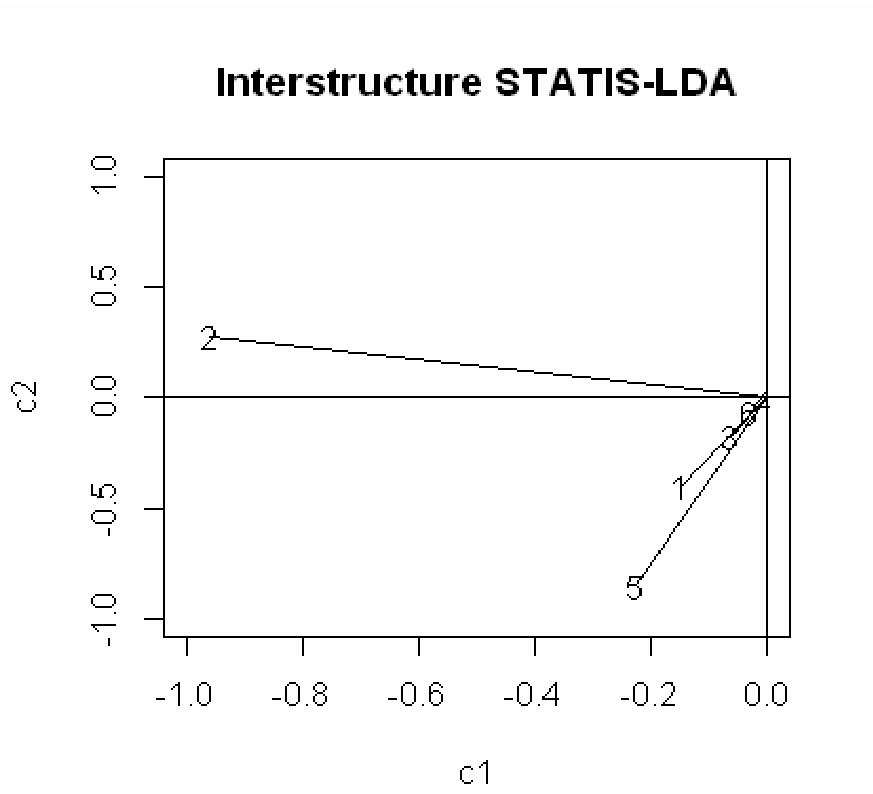
Tab:	1	2	3	4	5	6
gcr:	0.098	0.810	0.122	0.112	0.116	0.082

Global-LDA results:

Train gcr: 0.982
Test gcr: 0.744

Global-STATIS-LDA results:

Train gcr: 0.889
Test gcr: 0.778
Compromise coefficients:



.091 0.007 0.340

0.043

Conclusion and perspectives

- A very simple and rapid method to implement
- Uses classifical criteria of discrimination: between, within-variances,…
- Simultaneously provide global and individual results
- Gives easy insight in individual table importance in discrimination
  => An important interpretation tool
- Allows table weighting according to this importance
  => An better discrimination tool than joint-LDA when too noisy variables are involved
- Can be easily generalized to 4-four way tables,…
- Compromise coefficients seem to be useful quantification of individual table importance

Thank you for your attention