

Problème d'analyse des données corrigé

Graphiques et données

D. Chessel

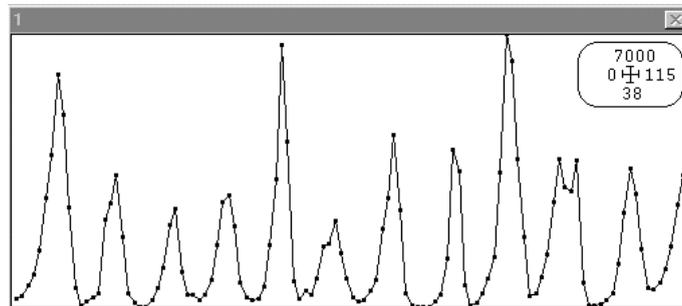
DEA AMSB / 1999 (Stage de 3 jours)

Questions

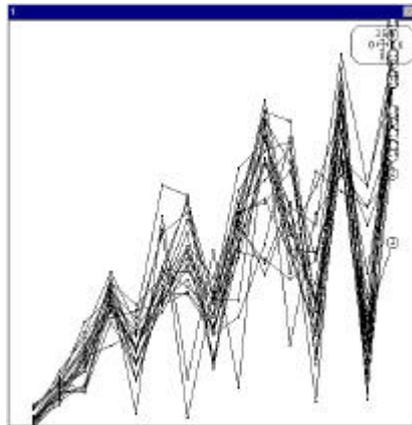
Chaque question porte sur une carte de données de la pile Data du logiciel ADE-4.

1 - Budget_E Que pensez-vous de la figure présente sur cette carte ? **ATTENTION** : il est inutile de créer le dossier de travail correspondant.

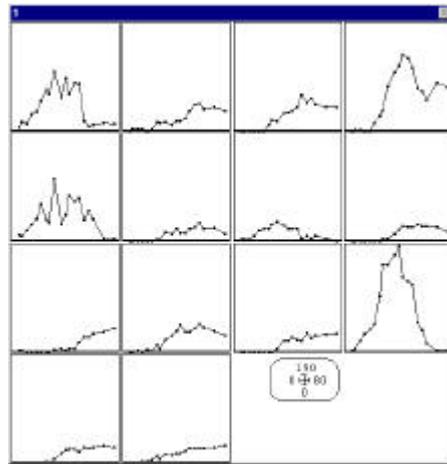
2 - Lynx Comment est obtenue cette figure et qu'exprime t'elle ?



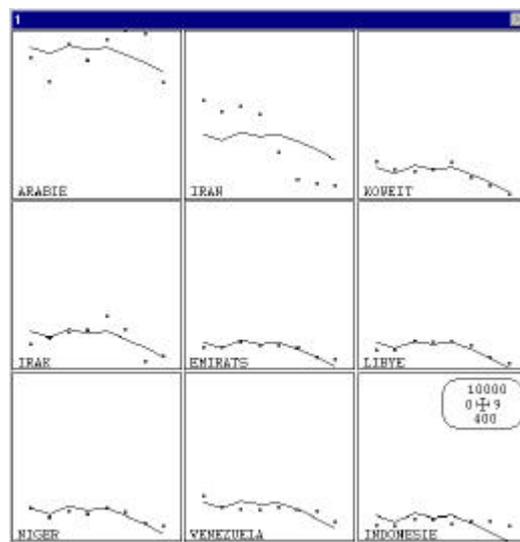
3 - Clementi Comment est obtenue cette figure et qu'exprime t'elle ?



4 - Grebes Comment est obtenue cette figure et qu'exprime t'elle ?



5 - **OPEP** Comment est obtenue cette figure et qu'exprime t'elle ?



6 - **Enquete** D'où proviennent les données reproduites dans cette carte et à quelle méthode renvoient t'elles ?

7 - **Tortues** Quel modèle simple utiliser pour le tableau Males ?

8 - **Rhone** Combien d'axes peut-on conserver dans l'ACP normée du tableau Rh ?

9 - **Wisconsin** Combien d'axes peut-on conserver dans l'AFC du tableau Wisconsin ?

A - **Bordeaux** Quelle relation existe t'elle entre la qualité des vins et le jugement des consommateurs ?

B - **EleParis** Caractériser le lien entre résultats aux élections et prix des appartements.

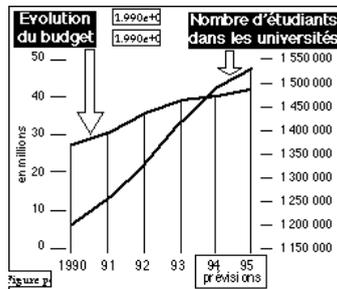
C - **Ombres** Pensez-vous qu'on puisse parler de variations de la forme des individus entre les populations d'Ombres communs décrites dans les données ?

D - **Sorme** Peut t'on utiliser l'analyse triadique partielle (STATIS: Table averaging), l'analyse STATIS (STATIS: Operator averaging), l'AFC de Foucart (STATIS: Foucart's COA) et/ou

l'analyse factorielle multiple (KTA-MFA: Multiple Factorial Analysis) sur le tableau de cette carte ? Si il y a plusieurs méthodes possibles, que doit-on faire ?

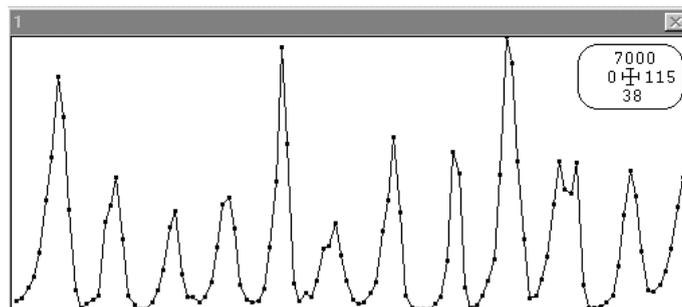
Réponses

Budget_E Que pensez-vous de la figure présente sur cette carte ?



En première lecture, elle fait apparaître que le budget de l'éducation nationale augmente moins vite que le nombre d'étudiants. En fait, cette représentation à deux échelles est une escroquerie intellectuelle, la même technique pouvant conduire à la lecture inverse.

Lynx Comment est obtenue cette figure et qu'exprime t'elle ?



Elle est tracée avec Curves: Lines avec le seul paramètre :

Y file (no default)

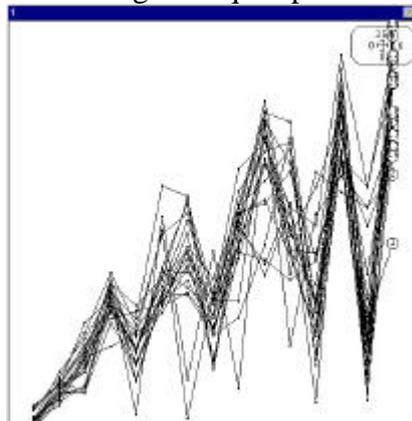
Set

E:\Ade4\LYNX\Lynx

114 1

Elle utilise une forme de la fenêtre physique adaptée (200-500 pixels). Elle montre le caractère périodique de la chronique.

Clementi Comment est obtenue cette figure et qu'exprime t'elle ?

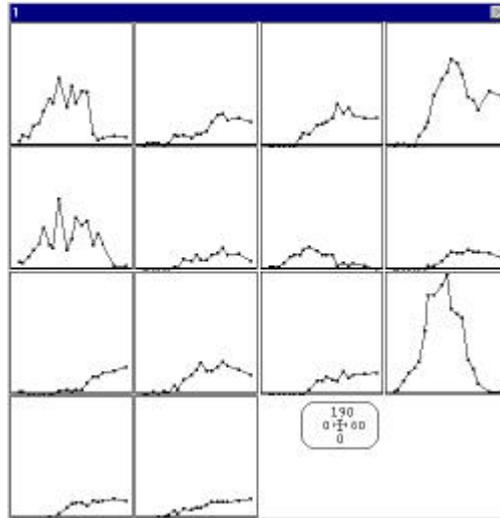


Elle est tracée avec Curves: Lines. On a du transposé le fichier de la carte de données pour mettre les dates en lignes et demandé :

Number of curves by window

Elle montre la croissance de la production des arbres et l'alternance d'une année sur l'autre. Toutes les chroniques ne sont pas totalement en phase.

Grebes Comment est obtenue cette figure et qu'exprime t'elle ?

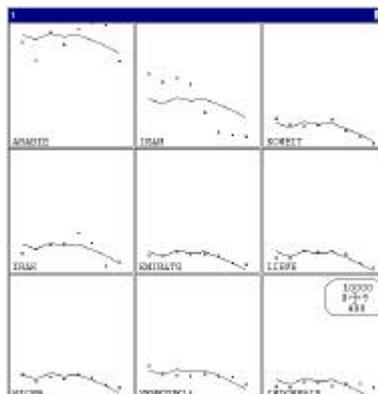


Elle est tracée avec Curves: Lines. On a demandé :

X file (default = 1, 2, 3, ..., n)	<input type="button" value="Set"/>	<input type="text" value="E:\Ade4\GREBES\date"/>	19	1
X file column number (default	<input type="button" value="Set"/>	<input type="text" value=""/>		
Y file (no default)	<input type="button" value="Set"/>	<input type="text" value="E:\Ade4\GREBES\Gre"/>	19	14

Les bornes de la fenêtre logique ont été modifiées. Elle montre la saisonnalité de la présence de grèbes sur les étangs et la variabilité des patterns de reproduction d'un étang à l'autre.

OPEP Comment est obtenue cette figure et qu'exprime t'elle ?



Elle est obtenue avec CurveModels: Numerical sur :

X file (default = 1, 2, 3, ..., n)	Set			
X file column number (default	Set			
Model values file (no default)	Set	E:\Ade4\OPEP\Opep.ccmo	8	9
Data values file (no default)	Set	E:\Ade4\OPEP\Opep	8	9
Variable label file (or #)	Set	E:\Ade4\OPEP\Label_Pays		

Le fichier Opep.ccmo est issu de HTA: Double centring additive sur Opep. Elle montre une tendance générale de la période à la limitation de la production (choc pétrolier) et les écarts à la tendance dans chaque pays.

Enquete D'où proviennent les données reproduites dans cette carte et à quelle méthode renvoient t'elles ?

La référence donnée est :

Rouanet, H. & Le Roux, B. (1993) Analyse des données multidimensionnelles. Dunod, Paris.

Elles demande une analyse des correspondances multiples à pondération non uniforme.

Tortues Quel modèle simple utiliser pour le tableau Males ?

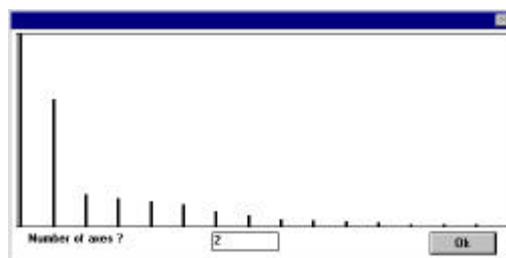
On utilise HTA: Edit inertia qui donne :

```
input file: E:\Ade4\TORTUES\Males
Option 1 = No action (non centred PCA)           Inertia = 32367.6
Option 2 = Centred table (overall centred PCA)   Inertia = 4285.9
Option 3 = Centred (zero mean) columns           Inertia = 666.038
Option 4 = Centred (zero mean) rows              Inertia = 3717
Option 5 = Additive model                        Inertia = 97.1424
Option 6 = Multiplicative model                  Inertia = 19.5128
```

Le modèle de base est du type $x_{ij} = k\mathbf{a}_i\mathbf{b}_j$.

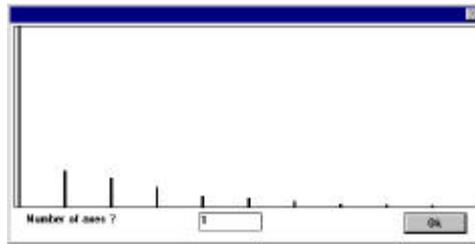
Rhone Combien d'axes peut-on conserver dans l'ACP normée du tableau Rh ?

On utilise PCA: Correlation matrix PCA et le graphe des valeurs propres montre deux axes :



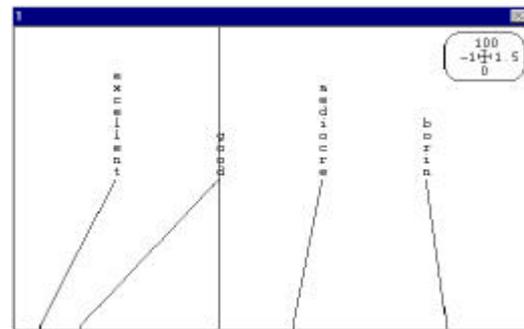
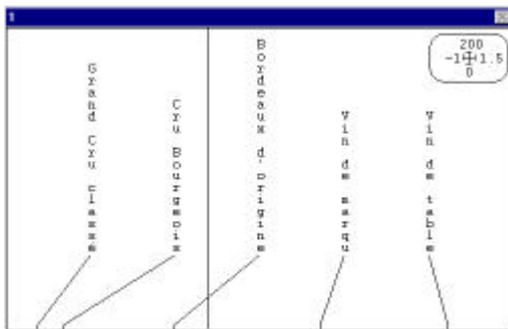
Wisconsin Combien d'axes peut-on conserver dans l'AFC du tableau Wisconsin ?

On utilise COA: COrrespondence Analysis et le graphe des valeurs propres montre un axe :



Bordeaux Quelle relation existe t'elle entre la qualité des vins et le jugement des consommateurs ?

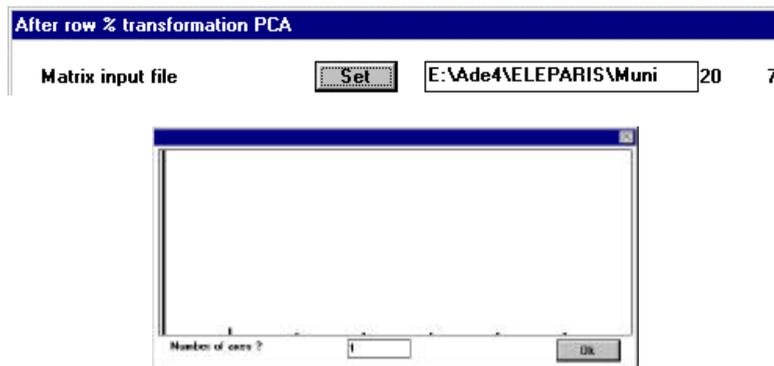
On fait l'analyse des correspondances du tableau (la corrélation canonique vaut 0.77) et on utilise (par exemple) Graph1D: Labels :



L'échelle des valeurs est respectée ! (N.B. Le dernier caractère de la dernière étiquette est absente par manque d'un RC dans les fichiers).

EleParis Caractérisez le lien entre résultats aux élections et prix des appartements.

On peut synthétiser le tableau des résultats par une ACP :

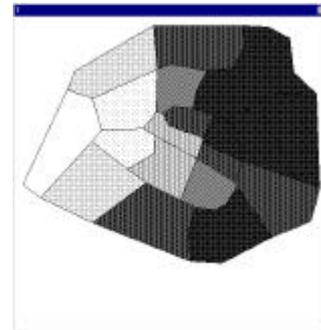
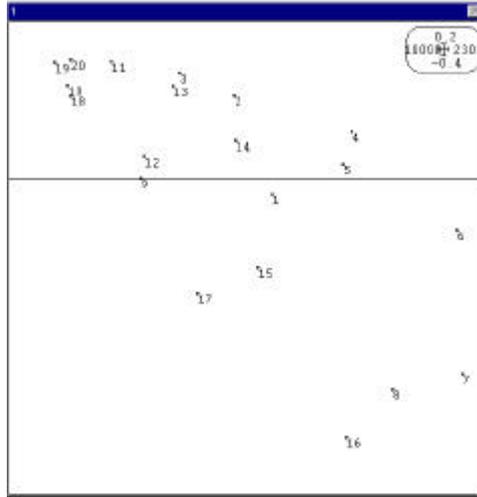
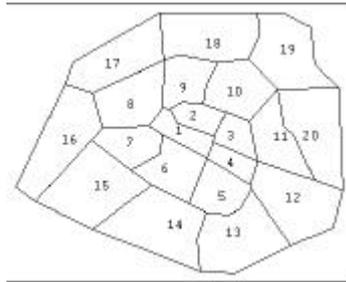


La corrélation entre score et prix vaut -0.71.

Diagonal inner product C-XDY

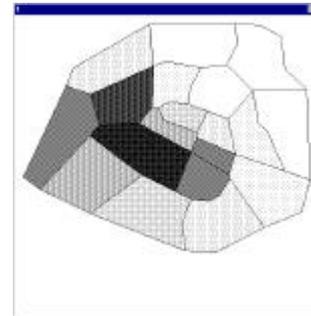
Input file for X matrix	Set	E:\Ade4\VELEPARIS\MuniRP	20	1
Option for X matrix	Set	3		
Input file for Y matrix	Set	E:\Ade4\VELEPARIS\Prix	20	1
Option for Y matrix	Set	3		

[1] -7.1597e-01



Gray levels areas

---.area type file	Set	E:\Ade4\VELEPARIS\Paris.are		
Data file	Set	Ade4\VELEPARIS\MuniRP.cpl	20	1

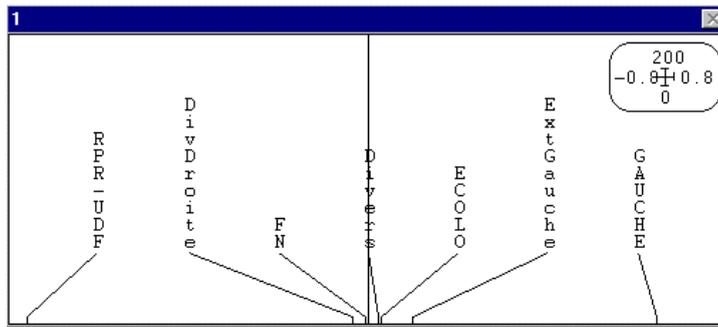


Gray levels areas

---.area type file	Set	E:\Ade4\VELEPARIS\Paris.are		
Data file	Set	E:\Ade4\VELEPARIS\Prix	20	1

Labels

Data file (no default)	Set	de4\VELEPARIS\MuniRP.cpc1	7	1
Rows label file (default = #)	Set	de4\VELEPARIS\Label_Candi		
Variable label file (or #)	Set			
Vertical (1) or horizontal (2)	Set	2		



C'est une affaire Droite (à gauche) contre Gauche (à droite). Moins on habite un quartier cher et plus on vote à Gauche. Et inversement. Avec un écart à la loi générale dans le 6^o arrondissement (les universitaires du quartier latin ?).

Ombres Pensez-vous qu'on puisse parler de variations de la forme des individus entre les populations d'Ombres communs décrites dans les données ?

On peut enlever l'effet taille par un modèle multiplicatif justifié par HTA: Edit inertia :

```
input file: E:\Ade4\OMBRES\Omb
Option 1 = No action (non centred PCA)           Inertia = 899833
Option 2 = Centred table (overall centred PCA)   Inertia = 316313
Option 3 = Centred (zero mean) columns           Inertia = 14712.6
Option 4 = Centred (zero mean) rows              Inertia = 308538
Option 5 = Additive model                        Inertia = 6937.78
Option 6 = Multiplicative model                  Inertia = 1928.59
```

On associe le tableau doublement centré à la variable qualitative Pop (Discrimin: Initialize/LinkPrep) :

Initialize: LinkPrep

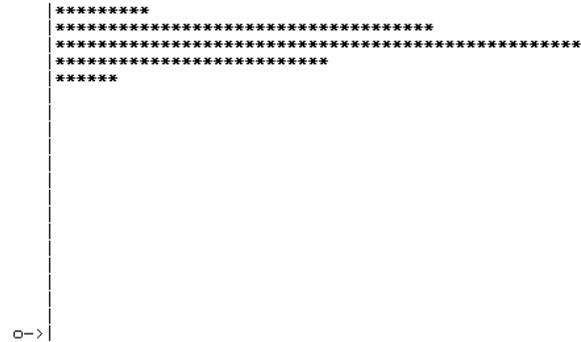
Statistical triplet	Set	E:\Ade4\OMBRES\Omb.dmta 120 13
Categories file (.cat)	Set	E:\Ade4\OMBRES\Pop.cat
Selected variable (default=1)	Set	
Output file name	Set	aa

La réponse est OUI car le test sur l'analyse discriminante est très significatif (Discrimin: Discriminant analysis/Test) :

Discriminant analysis: Test

Data input file	Set	E:\Ade4\OMBRES\aa.dis
Select a number of	Set	1000

```
number of random matching: 1000   Observed: 1.740782
Histogramm:  minimum = 0.227818, maximum = 1.740782
number of simulation X<Obs: 1000 (frequency: 1.000000)
number of simulation X>Obs: 0 (frequency: 0.000000)
```



On peut faire la même chose avec une ACP doublement centrée sur les log.

Sorme Peux t'on utiliser l'analyse triadique partielle (STATIS:Table averaging), l'analyse STATIS (STATIS:Operator averaging), l'AFC de Foucart (STATIS:Foucart's COA) et/ou l'analyse factorielle multiple (KTA-MFA: Multiple Factorial Analysis) sur le tableau de cette carte ? Si il y a plusieurs méthodes possibles, que doit-on faire ?

- STATIS: Table averaging | OUI avec 12 tableaux 10-10 complets
- STATIS: Operator averaging | OUI avec de deux manières différentes (WD ou VQ)
- STATIS: Foucart's COA | NON (variables quantitatives d'unités différentes)
- KTA-MFA: Multiple Factorial Analysis | OUI (mêmes individus stations)

Le plus simple est le meilleur (STATIS: Table averaging)