

Fiche de Biostatistique

Problèmes d'ACP

D. Chessel

Résumé

La fiche contient 7 problèmes sur l'analyse en composantes principales.

Plan

1.	TESTER UN PROGRAMME D'ACP.....	2
2.	SYNDICATS.....	3
3.	ACCIDENTS DE LA ROUTE	5
4.	BIBLIOTHEQUE	11
5.	DEGUSTATION DE BEAUJOLAIS.....	13
6.	TABLEAUX ARTIFICIELS	18
7.	EUROPE	25

1. Tester un programme d'ACP

On se propose, pour tester un programme, de faire l'analyse en composantes principales (option centrée) du tableau à 3 lignes (individus) et 3 colonnes (variables) :

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Toutes les réponses seront données sous forme mathématique sans approximation numérique.

- 1 Quelle est la trace de la matrice \mathbf{A} ? Quel est son déterminant?
- 2 Quel est le rang de \mathbf{A} ? Donner une base de son noyau?
- 3 Pourquoi est-elle diagonalisable?
- 4 Donner ses valeurs propres avec leur ordre de multiplicité.
- 5 Donner une base de vecteurs propres orthonormée.
- 6 \mathbb{R}^3 est muni du produit scalaire $\langle \mathbf{x} | \mathbf{y} \rangle = \sum_{i=1}^3 x_i y_i$. Soit f l'application de $\mathbb{R}^3 \times \mathbb{R}^3$ dans \mathbb{R} qui au couple (\mathbf{x}, \mathbf{y}) associe $f(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x} | \mathbf{A}\mathbf{y} \rangle$. L'application f est-elle un produit scalaire?
- 7 Calculer les valeurs propres de l'ACP centrée de \mathbf{A} en fonction des valeurs propres de \mathbf{A} .
- 8 Les vecteurs propres normés de \mathbf{A} et les axes principaux de l'ACP de \mathbf{A} sont-ils comparables ?
- 9 Le listing du programme testé donne pour l'ACP centrée de \mathbf{A} les résultats :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: E:\Oper\Sabine\CalADE\a.cpta
--- Number of rows: 3, columns: 3
-----
Total inertia: 3.33333
-----
Num. Eigenval.  R.Iner.  R.Sum      |Num. Eigenval.  R.Iner.  R.Sum |
01  +3.0000E+00 +0.9000 +0.9000  |02  +3.3333E-01 +0.1000 +1.0000 |
03  +0.0000E+00 +0.0000 +1.0000

File E:\Oper\Sabine\CalADE\a.cvpv contains the eigenvalues and relative inertia
It has 3 rows and 2 columns

File E:\Oper\Sabine\CalADE\a.cpcv contains the column scores
--- It has 3 rows and 2 columns
File :E:\Oper\Sabine\CalADE\a.cpcv
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-7.071e-01| 1.414e+00|
|  2|-4.082e-01| 4.082e-01|
|----|-----|-----|

File E:\Oper\Sabine\CalADE\a.cpli contains the row scores
--- It has 3 rows and 2 columns
File :E:\Oper\Sabine\CalADE\a.cpli
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-1.225e+00| 2.449e+00|
|  2|-7.071e-01| 7.071e-01|
```

|----|-----|-----|

Contrôler à l'aide des résultats des questions précédentes ces résultats numériques Le test est-il satisfaisant?

2. Syndicats

Une enquête d'opinion (période 1970-1980) portant sur 1000 personnes contient entre autres la question "A quelle famille politique vous rattachez vous ?" avec 5 modalités de réponse (EG : extrême gauche; G : gauche; C : centre, D : droite et ED : extrême droite) et la question " Que pensez vous de l'importance des syndicats ?" avec 4 modalités de réponse (1- importance trop grande; 2- importance convenable; 3- importance insuffisante; 4- sans opinion). La répartition des réponses est consignée dans la table de contingence :

	1	2	3	4
EG	10	31	85	4
G	80	90	118	52
C	55	58	31	36
D	95	38	46	71
ED	40	23	10	27

En considérant les familles politiques comme individus et les opinions à la question sur l'importance des syndicats comme variables on obtient la tableau 5-4 en pourcentage par lignes de l'annexe A. Les moyennes et les variances sont (pondération uniforme dans \mathbb{R}^5) dans l'annexe B. Le tableau centré, noté X, est consigné dans l'annexe C. On utilise la métrique identité dans \mathbb{R}^4 pour faire l'ACP du tableau X. La matrice de covariance figure dans l'annexe D.

III-1 Donner les quatre valeurs numériques qui ont été effacées dans les annexes A, B, C et D.

III-2 L'inertie totale est $6,447361 \cdot 10^{-2}$. Sachant que la matrice de covariance admet $0,3987 \cdot 10^{-2}$ et $0,2574 \cdot 10^{-4}$ comme valeurs propres donner la liste des valeurs propres de cette matrice de covariance et la répartition de l'inertie entre les axes principaux. Quel est le dernier vecteur propre ?

III-3 Donner la dimension, le rang et la liste des valeurs propres avec leur ordre de multiplicité de la matrice $(1/5)XX'$. Donner un vecteur du noyau.

II-4 Les coordonnées factorielles sont dans l'annexe E. Compléter et interpréter les cartes cartes factorielles.

Annexe A

```

0.077 0.238 0.654 0.031
0.235 0.265 0.347 0.153
0.306 0.322 0.172 0.200
0.380 0.152 0.184 0.284
0.400 0.230 0.100 xxxxxx
    
```

Annexe B

```

MOY:+.2796E+00   VAR:+.1366E-01
MOY:+.2415E+00   VAR:+.3041E-02
xxxxxxxxxxxxxxxxx VAR:+.3937E-01
MOY:+.1875E+00   VAR:+.8407E-02
    
```

Annexe C

```

xxxxxxx  -0.003  +0.362  -0.157
-0.044  +0.023  +0.056  -0.035
+0.026  +0.081  -0.119  +0.012
+0.100  -0.089  -0.107  +0.096
+0.120  -0.011  -0.191  +0.082
    
```

Annexe D

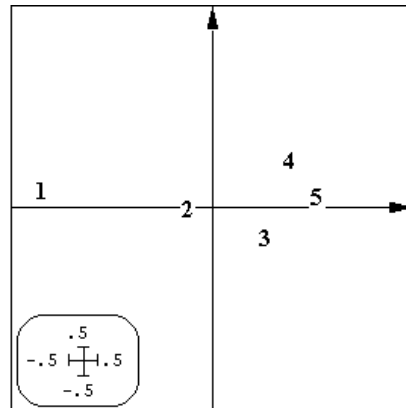
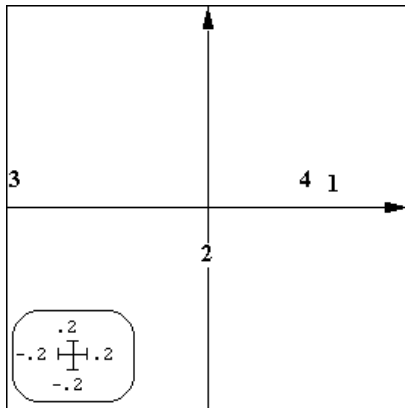
```

+.1366E-01  -.1738E-02  -.2257E-01  xxxxxxxxxxxx
-.1738E-02  +.3041E-02  +.4767E-03  -.1780E-02
-.2257E-01  +.4767E-03  +.3937E-01  -.1727E-01
+.1065E-01  -.1780E-02  -.1727E-01  +.8407E-02
    
```

Annexe E

```

xxxxxxxxxxxxx +.1990E-01
-.7886E-01  -.2383E-01
+.1099E+00  -.9732E-01
+.1716E+00  +.9738E-01
+.2408E+00  +.3870E-02
-----
+.1158E+00  +.1501E-01
-.7919E-02  -.5458E-01
xxxxxxxxxxxxx +.1991E-01
+.8948E-01  +.1965E-01
    
```



3. Accidents de la route

Une étude sur la morbidité par accidents de la route (1958-1960) donne en fonction de la classe d'âge la répartition des victimes par catégorie d'usagers. Il y a 5 catégories :

1-piétons, 2-cyclistes, 3-motocyclistes, 4-automobilistes, 5-autres usagers.

Il y a 14 classes d'âge exprimées en années :

1-[0-5[, 2-[5,10[, 3-[10,15[, etc ..., 14-[65-70[

Il y a deux tableaux. Le premier concernent les hommes :

FICHIER : HPC

```
-----  
Lig  1/0.552/0.011/0.035/0.375/0.027/  
Lig  2/0.630/0.096/0.028/0.224/0.022/  
Lig  3/0.247/0.378/0.141/0.201/0.033/  
Lig  4/0.045/0.124/0.681/0.126/0.024/  
Lig  5/0.046/0.045/0.611/0.250/0.048/  
Lig  6/0.050/0.051/0.537/0.305/0.057/  
Lig  7/0.063/0.057/0.466/0.347/0.067/  
Lig  8/0.069/0.064/0.440/0.358/0.069/  
Lig  9/0.086/0.079/0.440/0.335/0.060/  
Lig 10/0.105/0.096/0.434/0.304/0.061/  
Lig 11/0.128/0.119/0.415/0.283/0.055/  
Lig 12/0.154/0.141/0.419/0.270/0.016/  
Lig 13/0.201/0.160/0.339/0.256/0.044/  
Lig 14/0.346/0.182/0.221/0.219/0.032/  
-----
```

et le second concernent les femmes :

FICHIER : FPC

```
-----  
Lig  1/0.473/0.008/0.034/0.454/0.031/  
Lig  2/0.576/0.045/0.030/0.322/0.027/  
Lig  3/0.307/0.176/0.116/0.365/0.036/  
Lig  4/0.126/0.127/0.468/0.252/0.027/  
Lig  5/0.086/0.062/0.450/0.377/0.025/  
Lig  6/0.092/0.054/0.323/0.502/0.029/  
Lig  7/0.097/0.059/0.267/0.547/0.030/  
Lig  8/0.106/0.063/0.245/0.549/0.037/  
Lig  9/0.134/0.076/0.243/0.507/0.040/  
Lig 10/0.169/0.089/0.224/0.481/0.037/  
Lig 11/0.200/0.098/0.200/0.461/0.041/  
Lig 12/0.270/0.106/0.150/0.430/0.044/  
Lig 13/0.384/0.090/0.097/0.387/0.042/  
Lig 14/0.636/0.039/0.021/0.271/0.033/  
-----
```

On s'intéresse aux liaisons âge-type d'usage de la voie publique en ce qu'elles ont de commun et de spécifique du point de vue du sexe des victimes. Toutes les pondérations sur les lignes sont uniformes (1/n) et toutes les pondérations sur les colonnes sont unitaires (métriques canoniques).

I- ACP du tableau juxtaposé

On peut juxtaposer les tableaux HPC et FPC pour obtenir le tableau HF avec 14 lignes et 10 colonnes. Le listing de l'ACP centrée par colonnes (ACP classique sur matrice de covariance) est consigné dans l'annexe 1.

I-1 Donner le rang et une base de vecteurs propres orthonormés du noyau de la matrice de covariance.

I-2 En déduire les propriétés attendues des coordonnées factorielles.

I-3 Rédiger une légende pour la figure 1 après l'avoir complétée.

I-4 Dépouiller et interpréter cette analyse. Résumer le mode d'évolution de la répartition des types d'accidents en fonction de l'âge des victimes.

II- ACP du tableau superposé

On peut superposer les tableaux HPC et FPC pour obtenir le tableau H/F avec 28 lignes et 5 colonnes. Les principaux résultats de l'ACP centrée par colonnes (ACP classique sur matrice de covariance) de H/F sont consignés dans l'annexe 2.

II-1 Calculer les moyennes et les variances de cette analyse.

II-2 Quelle information exprimée est-elle commune à cette analyse et la précédente ?

II-3 Quelle information nouvelle est-elle exprimée ?

II-4 Rédiger une légende pour la figure 2 après l'avoir complétée.

III- ACP des différences

On veut centrer l'analyse sur les différences entre les deux tableaux de départ. Pour ce faire on considère le tableau des différences terme à terme HPC-FPC (14 lignes et 5 colonnes). On pratique sur ce tableau une ACP non centrée avec pondération uniforme des lignes et métrique canonique sur les colonnes.

III-1 Définir avec le schéma de dualité le critère optimisé par l'ACP non centrée de ce nouveau tableau. Justifier la présence d'une valeur propre nulle.

III-2 Interpréter les résultats consignés dans l'annexe 3.

III-3 Rédiger une légende pour la figure 3 après l'avoir complétée.

III-4 Résumer les objectifs poursuivis et les résultats obtenus dans ces trois types d'analyses sur un couple de tableaux. Connaissez-vous d'autres méthodes pour aborder des données du même type ?

Annexe 1

Fichier traite : HF

Nombre d'individus : 14 Nombre de variables : 10

MOYENNES-VARIANCES,

VAR: 1	MOY:+.1944E+00	VAR:+.3338E-01
VAR: 2	MOY:+.1145E+00	VAR:+.7446E-02
VAR: 3	MOY:+.3719E+00	VAR:+.3678E-01
VAR: 4	MOY:+.2752E+00	VAR:+.4422E-02
VAR: 5	MOY:+.4393E-01	VAR:+.3019E-03
VAR: 6	MOY:+.2611E+00	VAR:+.3277E-01
VAR: 7	MOY:+.7800E-01	VAR:+.1599E-02
VAR: 8	MOY:+.2049E+00	VAR:+.1905E-01

VAR: 9 MOY:+.4218E+00 VAR:+.8548E-02
 VAR: 10 MOY:+.3421E-01 VAR:+.3574E-04

INERTIE TOTALE .1443411

VALEURS PROPRES

1:.1121E+00/.7767/0.777 2:.1734E-01/.1201/0.897
 3:.9686E-02/.0671/0.964 4:.4520E-02/.0313/0.995
 5:.5027E-03/.0035/0.999 6:.1199E-03/.0008/1.000
 7:.5857E-04/.0004/1.000 8:.2250E-05/.0000/1.000

COORDONNEES DES LIGNES

< 1>+.5374E+00 -.1932E+00
 < 2>+.6560E+00 +.2552E-01
 < 3>+.2331E+00 +.1188E+00
 < 4>-.3989E+00 +.2917E+00
 < 5>-.3913E+00 +.7740E-01
 < 6>-.3102E+00 -.7035E-01
 < 7>-.2456E+00 -.1455E+00
 < 8>-.2153E+00 -.1578E+00
 < 9>-.1859E+00 -.1048E+00
 < 10>-.1439E+00 -.6073E-01
 < 11>-.9276E-01 -.2909E-01
 < 12>-.2091E-01 +.1988E-01
 < 13>+.1318E+00 +.5688E-01
 < 14>+.4464E+00 +.1714E+00

COORDONNEES DES COLONNES

< 1>+.1750E+00 -.7294E-02
 < 2>+.2265E-01 +.4796E-01
 < 3>-.1844E+00 +.3393E-01
 < 4>-.3427E-02 -.6486E-01
 < 5>-.9812E-02 -.9741E-02
 < 6>+.1719E+00 +.3904E-01
 < 7>-.1149E-01 +.2046E-01
 < 8>-.1265E+00 +.2510E-01
 < 9>-.3396E-01 -.8342E-01
 < 10>+.1421E-03 -.1178E-02

INERTIE DU NUAGE DES LIGNES

CONTRIBUTIONS ABSOLUES CONTRIBUTIONS RELATIVES

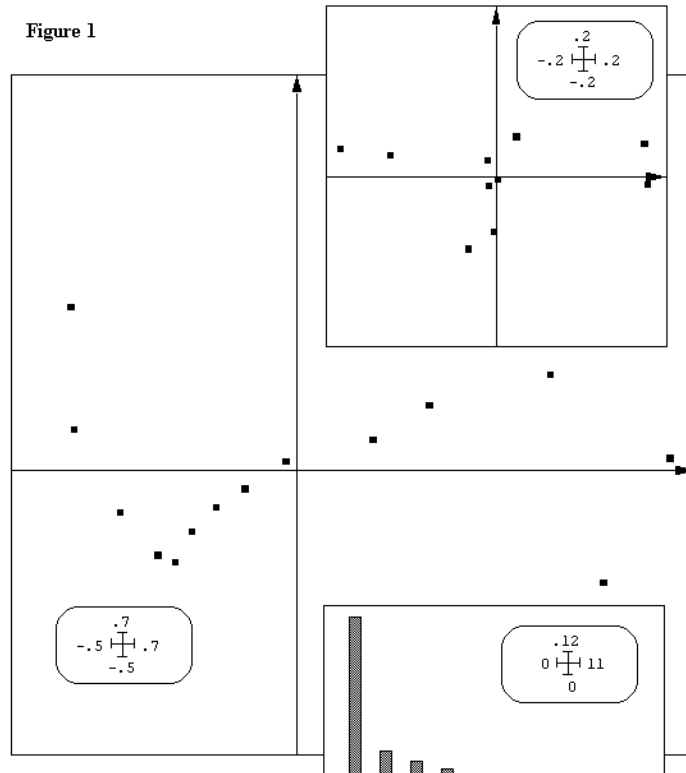
(NUM)	(FAC 1)	(FAC 2)	(NUM)	(FAC 1)	(FAC 2)	(RESTE)	(POIDS)	(C.TRA)
(1)	(1840)	(1538)	(1)	(8437)	(1091)	(473)	(714)	(1694)
(2)	(2742)	(27)	(2)	(9515)	(14)	(470)	(714)	(2238)
(3)	(346)	(581)	(3)	(3529)	(916)	(5555)	(714)	(762)
(4)	(1014)	(3505)	(4)	(6134)	(3280)	(586)	(714)	(1284)
(5)	(976)	(247)	(5)	(8613)	(337)	(1050)	(714)	(880)
(6)	(613)	(204)	(6)	(9360)	(481)	(159)	(714)	(509)
(7)	(384)	(873)	(7)	(7359)	(2584)	(57)	(714)	(406)
(8)	(295)	(1026)	(8)	(6391)	(3436)	(173)	(714)	(359)
(9)	(220)	(452)	(9)	(7448)	(2366)	(185)	(714)	(230)
(10)	(132)	(152)	(10)	(8025)	(1429)	(546)	(714)	(128)
(11)	(55)	(35)	(11)	(7043)	(692)	(2265)	(714)	(60)
(12)	(3)	(16)	(12)	(465)	(421)	(9114)	(714)	(46)
(13)	(111)	(133)	(13)	(5478)	(1020)	(3502)	(714)	(157)
(14)	(1270)	(1211)	(14)	(7901)	(1165)	(933)	(714)	(1248)

INERTIE DU NUAGE DES COLONNES

CONTRIBUTIONS ABSOLUES CONTRIBUTIONS RELATIVES

(NUM)	(FAC 1)	(FAC 2)	(NUM)	(FAC 1)	(FAC 2)	(RESTE)	(POIDS)	(C.TRA)
(1)	(2733)	(31)	(1)	(9177)	(16)	(807)	(10000)	(2313)
(2)	(46)	(1327)	(2)	(689)	(3089)	(6222)	(10000)	(516)
(3)	(3034)	(664)	(3)	(9249)	(313)	(439)	(10000)	(2548)
(4)	(1)	(2426)	(4)	(27)	(9513)	(460)	(10000)	(306)
(5)	(9)	(55)	(5)	(3189)	(3143)	(3668)	(10000)	(21)
(6)	(2634)	(879)	(6)	(9013)	(465)	(522)	(10000)	(2270)
(7)	(12)	(241)	(7)	(825)	(2617)	(6558)	(10000)	(111)
(8)	(1428)	(363)	(8)	(8406)	(331)	(1263)	(10000)	(1320)
(9)	(103)	(4013)	(9)	(1350)	(8141)	(509)	(10000)	(592)
(10)	(0)	(1)	(10)	(6)	(388)	(9606)	(10000)	(2)

TOUTES LES CONTRIBUTIONS SONT EXPRIMEES EN 1/10000



Annexe 2

INERTIE TOTALE 8.598892E-02
 TABLEAU CENTRE

```

-----
< 1>+.3242E+00 -.8525E-01 -.2534E+00 +.2650E-01 -.1207E-01
< 2>+.4022E+00 -.2500E-03 -.2604E+00 -.1245E+00 -.1707E-01
< 3>+.1921E-01 +.2818E+00 -.1474E+00 -.1475E+00 -.6071E-02
< 4>-.1828E+00 +.2775E-01 +.3926E+00 -.2225E+00 -.1507E-01
< 5>-.1818E+00 -.5125E-01 +.3226E+00 -.9850E-01 +.8929E-02
< 6>-.1778E+00 -.4525E-01 +.2486E+00 -.4350E-01 +.1793E-01
< 7>-.1648E+00 -.3925E-01 +.1776E+00 -.1500E-02 +.2793E-01
< 8>-.1588E+00 -.3225E-01 +.1516E+00 +.9500E-02 +.2993E-01
< 9>-.1418E+00 -.1725E-01 +.1516E+00 -.1350E-01 +.2093E-01
< 10>-.1228E+00 -.2500E-03 +.1456E+00 -.4450E-01 +.2193E-01
< 11>-.9979E-01 +.2275E-01 +.1266E+00 -.6550E-01 +.1593E-01
< 12>-.7379E-01 +.4475E-01 +.1306E+00 -.7850E-01 -.2307E-01
< 13>-.2679E-01 +.6375E-01 +.5061E-01 -.9250E-01 +.4929E-02
< 14>+.1182E+00 +.8575E-01 -.6739E-01 -.1295E+00 -.7071E-02
< 15>+.2452E+00 -.8825E-01 -.2544E+00 +.1055E+00 -.8071E-02
< 16>+.3482E+00 -.5125E-01 -.2584E+00 -.2650E-01 -.1207E-01
< 17>+.7921E-01 +.7975E-01 -.1724E+00 +.1650E-01 -.3071E-02
< 18>-.1018E+00 +.3075E-01 +.1796E+00 -.9650E-01 -.1207E-01
< 19>-.1418E+00 -.3425E-01 +.1616E+00 +.2850E-01 -.1407E-01
< 20>-.1358E+00 -.4225E-01 +.3461E-01 +.1535E+00 -.1007E-01
< 21>-.1308E+00 -.3725E-01 -.2139E-01 +.1985E+00 -.9071E-02
< 22>-.1218E+00 -.3325E-01 -.4339E-01 +.2005E+00 -.2071E-02
< 23>-.9379E-01 -.2025E-01 -.4539E-01 +.1585E+00 +.9286E-03
< 24>-.5879E-01 -.7250E-02 -.6439E-01 +.1325E+00 -.2071E-02
< 25>-.2779E-01 +.1750E-02 -.8839E-01 +.1125E+00 +.1929E-02
< 26>+.4221E-01 +.9750E-02 -.1384E+00 +.8150E-01 +.4929E-02
< 27>+.1562E+00 -.6250E-02 -.1914E+00 +.3850E-01 +.2929E-02
< 28>+.4082E+00 -.5725E-01 -.2674E+00 -.7750E-01 -.6071E-02
    
```


VALEURS PROPRES

COORDONNEES DES COLONNES

```

-----
< 1>+.6425E-01 +.7472E+00 < 1>+.1770E+00 -.4543E-01
< 2>+.1581E-01 +.1838E+00 < 2>-.3444E-02 -.2809E-01
< 3>+.5733E-02 +.6667E-01 < 3>-.1809E+00 -.3613E-01
< 4>+.1947E-03 +.2264E-02 < 4>+.1313E-01 +.1079E+00
< 5>+.6186E-08 +.7193E-07 < 5>-.5755E-02 +.1736E-02

```

COORDONNEES DES LIGNES

```

-----
< 1>+.4100E+00 -.2707E-02 < 15>+.3596E+00 +.9466E-01
< 2>+.4606E+00 -.1776E+00 < 16>+.4271E+00 -.6303E-01
< 3>+.1073E+00 -.1542E+00 < 17>+.1782E+00 +.1722E-01
< 4>-.4193E+00 -.2442E+00 < 18>-.2044E+00 -.1047E+00
< 5>-.3617E+00 -.9999E-01 < 19>-.2120E+00 +.3671E-01
< 6>-.3036E+00 -.3418E-01 < 20>-.1107E+00 +.1802E+00
< 7>-.2420E+00 +.1637E-01 < 21>-.6504E-01 +.2320E+00
< 8>-.2188E+00 +.2958E-01 < 22>-.4317E-01 +.2360E+00
< 9>-.2081E+00 +.2218E-03 < 23>-.2462E-01 +.1875E+00
< 10>-.1924E+00 -.3531E-01 < 24>+.1192E-01 +.1551E+00
< 11>-.1641E+00 -.6141E-01 < 25>+.4944E-01 +.1316E+00
< 12>-.1489E+00 -.8857E-01 < 26>+.1322E+00 +.9236E-01
< 13>-.6058E-01 -.9844E-01 < 27>+.2476E+00 +.3304E-01
< 14>+.1229E+00 -.1538E+00 < 28>+.4727E+00 -.1245E+00

```

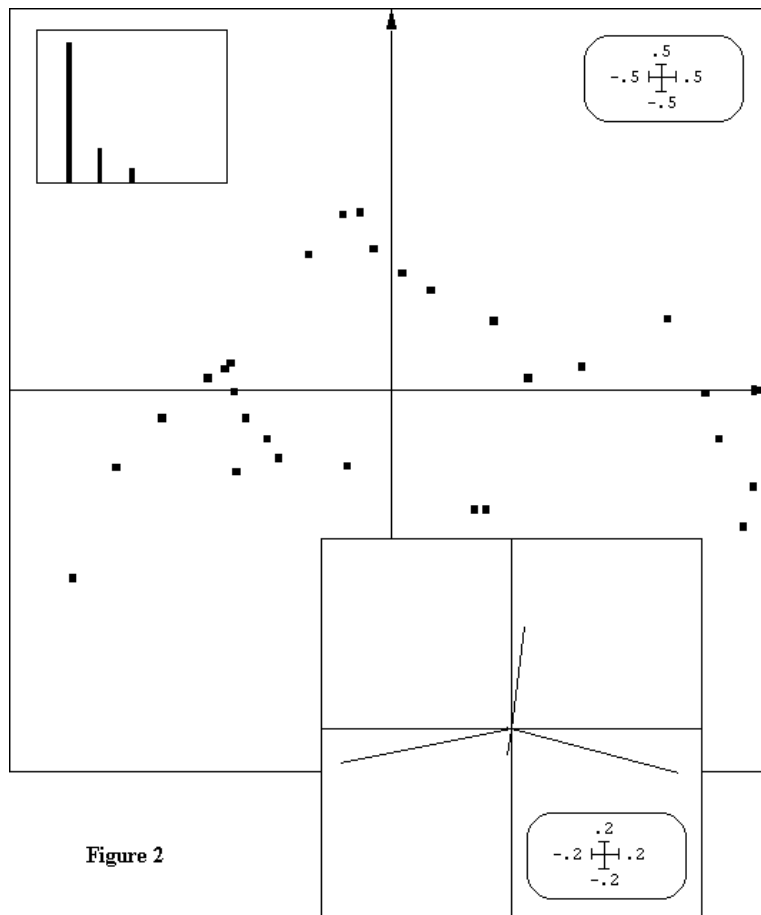


Figure 2

Annexe 3

TABLEAU TRAITE

```

< 1>+.7900E-01 +.3000E-02 +.1000E-02 -.7900E-01 -.4000E-02
< 2>+.5400E-01 +.5100E-01 -.2000E-02 -.9800E-01 -.5000E-02
< 3>-.6000E-01 +.2020E+00 +.2500E-01 -.1640E+00 -.3000E-02
< 4>-.8100E-01 -.3000E-02 +.2130E+00 -.1260E+00 -.3000E-02

```

```

< 5>-.4000E-01 -.1700E-01 +.1610E+00 -.1270E+00 +.2300E-01
< 6>-.4200E-01 -.3000E-02 +.2140E+00 -.1970E+00 +.2800E-01
< 7>-.3400E-01 -.2000E-02 +.1990E+00 -.2000E+00 +.3700E-01
< 8>-.3700E-01 +.1000E-02 +.1950E+00 -.1910E+00 +.3200E-01
< 9>-.4800E-01 +.3000E-02 +.1970E+00 -.1720E+00 +.2000E-01
< 10>-.6400E-01 +.7000E-02 +.2100E+00 -.1770E+00 +.2400E-01
< 11>-.7200E-01 +.2100E-01 +.2150E+00 -.1780E+00 +.1400E-01
< 12>-.1160E+00 +.3500E-01 +.2690E+00 -.1600E+00 -.2800E-01
< 13>-.1830E+00 +.7000E-01 +.2420E+00 -.1310E+00 +.2000E-02
< 14>-.2900E+00 +.1430E+00 +.2000E+00 -.5200E-01 -.1000E-02
    
```

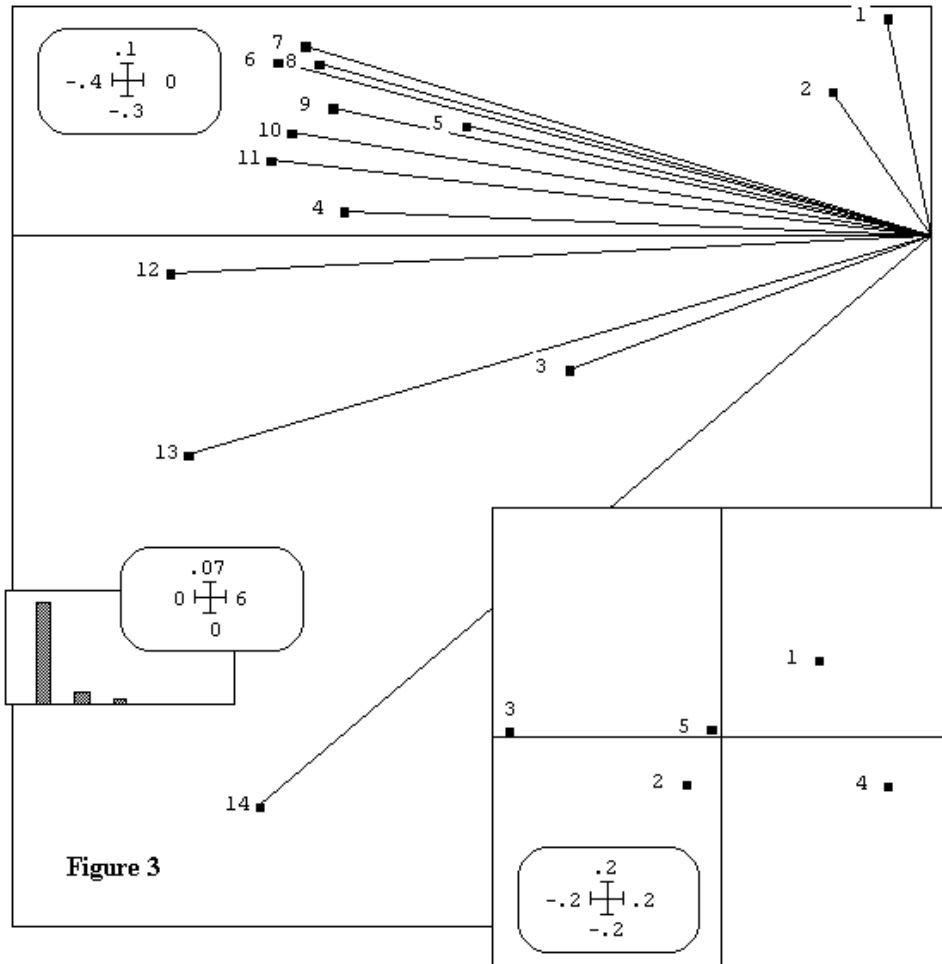


Figure 3

```

INERTIE TOTALE 7.627571E-02
VALEURS PROPRES
/ 1:.6359E-01/.8336/0.8336/ 2:.8125E-02/.1065/0.9402
/ 3:.4287E-02/.0562/0.9964/ 4:.2770E-03/.0036/1.0000
/ 5:.0000E+00/.0000/1.0000
COORDONNEES DES COLONNES
Lig 1/+.8480E-01 +.6809E-01
Lig 2/-.3220E-01 -.4034E-01
Lig 3/-.1860E+00 +.5154E-02
Lig 4/+.1438E+00 -.4188E-01
Lig 5/-.1042E-01 +.8970E-02
COORDONNEES DES LIGNES
Lig 1/-.1943E-01 +.9470E-01
Lig 2/-.4255E-01 +.6289E-01
Lig 3/-.1578E+00 -.5840E-01
Lig 4/-.2556E+00 +.1058E-01
Lig 5/-.2034E+00 +.4789E-01
Lig 6/-.2850E+00 +.7617E-01
Lig 7/-.2735E+00 +.8320E-01
Lig 8/-.2666E+00 +.7468E-01
Lig 9/-.2607E+00 +.5557E-01
Lig 10/-.2792E+00 +.4515E-01
    
```

Lig 11/-.2875E+00 +.3260E-01
 Lig 12/-.3319E+00 -.1636E-01
 Lig 13/-.3237E+00 -.9467E-01
 Lig 14/-.2929E+00 -.2476E+00

4. Bibliothèque

Les inscrits à la bibliothèque municipale de Lyon Part-Dieu entre 1979 et 1986 se répartissent par classe d'âge de la manière suivante :

		âge des inscrits en années					
		[... ,18[[18,25[[25,35[[35,50[[50,65[[65,...[
1979	*	752	36069	3413	1604	884	584
1980	*	1164	69331	6169	2594	1355	858
1981	*	1261	710327	7027	2865	1373	757
1982	*	1315	113834	10391	4094	1889	943
1983	*	2134	712190	7214	3054	1465	596
1984	*	2312	813422	8180	3602	1574	661
1985	*	2330	813592	8454	3812	1576	703
1986	*	1979	812588	8269	3725	1498	695

L'analyse en composantes principales de ces données donnent les résultats suivants.

TABLEAU D'ENTREE

```
1* 0.057 0.456 0.257 0.121 0.066 0.044
2* 0.054 0.435 0.287 0.121 0.063 0.040
3* 0.053 0.437 0.298 0.121 0.058 0.032
4* 0.041 0.426 0.320 0.126 0.058 0.029
5* 0.080 0.457 0.271 0.115 0.055 0.022
6* 0.078 0.451 0.275 0.121 0.053 0.022
7* 0.076 0.446 0.277 0.125 0.052 0.023
8* 0.069 0.438 0.288 0.130 0.052 0.024
```

MOYENNES ET VARIANCES DU FICHIER DE DEPART

```
1: +.6347E-01 .1799E-03
2: +.4433E+00 .1085E-03
3: +.2840E+00 .3210E-03
4: +.1224E+00 .1776E-04
5: +.5720E-01 .2519E-04
6: +.2960E-01 .6211E-04
```

TABLEAU CENTRE

```
1* -0.0069 0.0128 -0.0275 -0.0018 0.0092 0.0143
2* -0.0093 -0.0087 0.0033 -0.0016 0.0059 0.0104
3* -0.0101 -0.0059 0.0136 -0.0010 0.0010 0.0025
4* -0.0230 -0.0172 0.0360 0.0037 0.0010 -0.0006
5* 0.0166 0.0140 -0.0134 -0.0078 -0.0022 -0.0072
6* 0.0142 0.0078 -0.0091 -0.0013 -0.0043 -0.0074
7* 0.0130 0.0028 -0.0065 0.0027 -0.0055 -0.0065
8* 0.0054 -0.0055 0.0036 0.0072 -0.0051 -0.0054
```

DIAGONALISATION

1: .5213E-03/0.72970/0.72970 2: .1659E-03/0.23228/0.96198
 3: .1964E-04/0.02748/0.98946 4: .7293E-05/0.01021/0.99967
 5: .2328E-06/0.00033/1.00000 6: .0000E+00/0.00000/1.00000

COORDONNEES DES LIGNES			COORDONNEES DES COLONNES		
*	1	2	*	1	2
*	-----	-----	*	-----	-----
1*	0.0222	0.0275	1*	0.0110	-0.0076
2*	-0.0113	0.0122	2*	0.0100	0.0009
3*	-0.0177	0.0023	3*	-0.0172	-0.0049
4*	-0.0460	-0.0015	4*	-0.0021	-0.0010
5*	0.0253	-0.0083	5*	-0.0006	0.0049
6*	0.0177	-0.0104	6*	-0.0011	0.0077
7*	0.0126	-0.0112			
8*	-0.0028	-0.0106			

ANALYSE D'INERTIE

INERTIE TOTALE : 7.14437E-04

TOUTES LES CONTRIBUTIONS SONT EXPRIMEES EN 1/10000

***** CONTRIBUTIONS ABSOLUES DES LIGNES *****

[NUM]	[FAC 1]	[FAC 2]
[1]	[1182]	[5711]
[2]	[304]	[1116]
[3]	[752]	[39]
[4]	[5080]	[18]
[5]	[1533]	[520]
[6]	[750]	[807]
[7]	[381]	[941]
[8]	[19]	[849]

***** CONTRIBUTIONS RELATIVES DES LIGNES *****

[NUM]	[FAC 1]	[FAC 2]	[RESTE]	[POIDS]	[C.TRA]
[1]	[3905]	[6007]	[88]	[1250]	[2208]
[2]	[3989]	[4662]	[1350]	[1250]	[556]
[3]	[9523]	[159]	[318]	[1250]	[576]
[4]	[9912]	[11]	[77]	[1250]	[3740]
[5]	[8313]	[898]	[789]	[1250]	[1346]
[6]	[7426]	[2544]	[29]	[1250]	[737]
[7]	[5302]	[4167]	[532]	[1250]	[525]
[8]	[433]	[6302]	[3264]	[1250]	[313]

***** CONTRIBUTIONS ABSOLUES DES COLONNES *****

[NUM]	[FAC 1]	[FAC 2]
[1]	[2313]	[3446]
[2]	[1912]	[45]
[3]	[5662]	[1428]
[4]	[84]	[64]
[5]	[7]	[1448]
[6]	[23]	[3569]

***** CONTRIBUTIONS RELATIVES DES COLONNES *****

[NUM]	[FAC 1]	[FAC 2]	[RESTE]	[POIDS]	[C.TRA]
[1]	[6704]	[3180]	[116]	[10000]	[2517]
[2]	[9185]	[68]	[747]	[10000]	[1519]
[3]	[9195]	[738]	[67]	[10000]	[4493]
[4]	[2459]	[595]	[6946]	[10000]	[249]
[5]	[146]	[9541]	[312]	[10000]	[353]
[6]	[190]	[9537]	[273]	[10000]	[869]

Dépouiller cette analyse et caractériser simplement l'évolution de la distribution des âges des inscrits à la bibliothèque.

5. Dégustation de Beaujolais

DEA AMSB / Tronc commun / Décembre 1995

Les résultats forment un tableau Macon à 8 lignes (juges) et 25 colonnes (produits) :

```
5543347213544548578546728
4824152788163785781415446
2611621554372262162121251
6758268866656636817674167
1432716431281113226282812
3286583347815874433868673
7165474175738327354737385
8377835622427451645353534
```

On peut transposer le tableau : on appelle MTR le fichier qui contient 25 lignes (juges) et 8 colonnes (produits) :

```
54261378
58674213
42153867
34182657
31627548
45261873
72186345
27584316
18563472
38461752
51362874
46758132
43261587
57261834
48631725
85263471
57182436
78612354
81276345
54162873
41278635
65142873
74218635
24561783
86172354
```

Les dégustateurs sont de 5 catégories professionnelles. 1 à 5 sont oenologues, 6 à 10 sont restaurateurs, 11 à 15 sont négociants, 16 à 20 sont viticulteurs et 21 à 25 sont les organisateurs du concours.

On peut utiliser deux méthodes de sélection pour l'attribution des médailles :

A- Seront sélectionnés les produits ayant obtenu les meilleures sommes des rangs;

B- Sera d'abord sélectionné le produit ayant obtenu le plus grand nombre de places de premier, puis en cas d'ex æquo le plus grand nombre de places de second, puis en cas d'ex æquo le plus grand nombre de places de troisième, puis...

Les sommes des rangs par produit sont :

119 119 76 141 81 130 122 112

On entreprend sur le tableau de données initial, Macon, une analyse en composantes principales normée à pondération uniforme. On obtient le listing :

```

Classical Principal Component Analysis (Hotelling 1933)
Input file: Macon
---- Row weights:
File Macon.cnpl contains the row weights
It has 8 rows and 1 column
Each row has 1.2500e-01 weight (Sum = 1)
---- Column weights:
File Macon.cnpc contains the column weights
It has 8 rows and 1 column
Each column has unit weight (Sum = 25)
---- Table:
File Macon.cnta contains the centred and normed table
Zero mean and unit variance for each column
It has 8 rows and 25 columns
File :Macon.cnta
| Col. | Mini | Maxi |
|----|-----|-----|
| 1 | -1.528e+00 | 1.528e+00 |
| 2 | -1.528e+00 | 1.528e+00 |
...
| 24 | -1.528e+00 | 1.528e+00 |
| 25 | -1.528e+00 | 1.528e+00 |
|----|-----|-----|
---- Info: means and variances
File Macon.cnma contains the descriptive of the analysis
It contains successively:
    Number of rows: 8
    Number of columns: 25
    means and variances:
    Col.: 1 | Mean: 4.5000e+00 | Variance: 5.2500e+00
    Col.: 2 | Mean: 4.5000e+00 | Variance: 5.2500e+00
...
    Col.: 24 | Mean: 4.5000e+00 | Variance: 5.2500e+00
    Col.: 25 | Mean: 4.5000e+00 | Variance: 5.2500e+00
-----
File Macon.cn+r contains the Correlation matrix
from statistical triplet Macon.cnta
It has 25 rows and 25 columns
----- Correlation matrix -----

[ 1] 1000
[ 2] -238 1000
[ 3] 548 -714 1000
[ 4] 714 -167 762 1000
[ 5] -119 -548 214 -167 1000
[ 6] 381 -286 619 619 -524 1000
[ 7] 333 0 286 405 -95 -24 1000
[ 8] 48 738 -262 333 -190 -190 48 1000
[ 9] 24 286 -238 119 -571 429 -429 286 1000
[10] 48 286 71 333 -738 762 -381 238 738 1000
[11] 381 -619 786 548 -48 762 238 -452 -71 238 1000
[12] -619 619 -905 -667 -71 -690 -48 310 190 -238 -762
      1000
[13] 905 -500 762 786 -71 643 238 -143 119 190 690
      -786 1000

```

```
[ 14] 214 190 405 571 -595 738 -24 238 214 786 357
      -548 286 1000
[ 15] -95 333 -71 24 -238 238 -667 310 214 619 -143
      -238 -119 643 1000
[ 16] 238 48 48 95 -762 524 405 -381 190 333 381
      -119 262 310 -238 1000
[ 17] 571 429 262 738 -571 381 429 595 167 429 48
      -286 405 690 238 286 1000
[ 18] 24 262 -452 -476 -333 -71 -524 -190 143 214 -381
      71 -167 95 524 214 -48 1000
[ 19] 262 -95 238 214 48 -143 952 -143 -619 -548 262
      -48 167 -167 -690 381 214 -429 1000
[ 20] 381 -405 714 619 -452 952 214 -333 214 571 833
      -714 643 667 24 619 357 -190 119 1000
[ 21] -119 -262 429 333 357 -71 690 48 -452 -405 286
      -48 0 -95 -571 -119 48 -905 643 143 1000
[ 22] 310 -429 619 405 -452 881 48 -500 143 548 690
      -714 524 667 167 643 262 119 -0 929 -71 1000
[ 23] -333 -333 143 -310 357 -286 190 -452 -667 -476 -143
      -48 -357 -95 -143 -48 -238 24 286 -71 357 143
      1000
[ 24] 286 -310 405 429 -286 810 -310 -143 619 667 690
      -476 595 357 167 238 24 -119 -381 667 -214 524
      -643 1000
[ 25] 571 286 143 429 -738 405 571 71 48 286 190
      -238 405 500 -48 786 762 214 476 476 -95 476
      -95 0 1000
```

```
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: Macon.cnta
--- Number of rows: 8, columns: 25
```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+8.7740E+00	+0.3510	+0.3510	02	+5.6353E+00	+0.2254	+0.5764
03	+3.7414E+00	+0.1497	+0.7260	04	+2.6722E+00	+0.1069	+0.8329
05	+1.9532E+00	+0.0781	+0.9110	06	+1.7209E+00	+0.0688	+0.9799
07	+5.0310E-01	+0.0201	+1.0000	08	+0.0000E+00	+0.0000	+1.0000

```
File Macon.cnvp contains the eigenvalues and relative inertia for each axis
--- It has 25 rows and 2 columns
```

```
File Macon.cnco contains the column scores
--- It has 25 rows and 3 columns
```

```
File :Macon.cnco
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-7.918e-01| 9.414e-01|
|  2|-7.730e-01| 7.432e-01|
|  3|-4.838e-01| 7.031e-01|
|----|-----|-----|
```

```
File Macon.cnli contains the row scores
--- It has 8 rows and 3 columns
```

```
File :Macon.cnli
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-4.877e+00| 3.368e+00|
|  2|-2.816e+00| 5.044e+00|
|  3|-2.710e+00| 3.339e+00|
|----|-----|-----|
```

On entreprend sur le tableau transposé MTR une analyse en composantes principales normée à pondération uniforme. On obtient le listing :

```

Classical Principal Component Analysis (Hotelling 1933)
Input file: MTR
---- Row weights:
File MTR.cnpl contains the row weights
It has 25 rows and 1 column
Each row has 4.0000e-02 weight (Sum = 1)
---- Column weights:
File MTR.cnpc contains the column weights
It has 25 rows and 1 column
Each column has unit weight (Sum = 8)
---- Table:
File MTR.cnta contains the centred and normed table
Zero mean and unit variance for each column
It has 25 rows and 8 columns
File :MTR.cnta
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-1.976e+00| 1.702e+00|
|  2|-1.542e+00| 1.329e+00|
|  3|-1.015e+00| 1.971e+00|
|  4|-2.346e+00| 1.193e+00|
|  5|-9.315e-01| 1.979e+00|
|  6|-1.917e+00| 1.278e+00|
|  7|-1.844e+00| 1.483e+00|
|  8|-1.812e+00| 1.833e+00|
|----|-----|-----|
---- Info: means and variances
File MTR.cnma contains the descriptive of the analysis
It contains successively:
  Number of rows: 25
  Number of columns: 8
  means and variances:
  Col.:  1 | Mean:  4.7600e+00 | Variance: 3.6224e+00
  Col.:  2 | Mean:  4.7600e+00 | Variance: 5.9424e+00
  Col.:  3 | Mean:  3.0400e+00 | Variance: 4.0384e+00
  Col.:  4 | Mean:  5.6400e+00 | Variance: 3.9104e+00
  Col.:  5 | Mean:  3.2400e+00 | Variance: 5.7824e+00
  Col.:  6 | Mean:  5.2000e+00 | Variance: 4.8000e+00
  Col.:  7 | Mean:  4.8800e+00 | Variance: 4.4256e+00
  Col.:  8 | Mean:  4.4800e+00 | Variance: 3.6896e+00
-----
File MTR.cn+r contains the Correlation matrix
from statistical triplet MTR.cnta
It has 8 rows and 8 columns
----- Correlation matrix -----
[ 1] 1000
[ 2] -159 1000
[ 3] -416  410 1000
[ 4] -129  -59 -379 1000
[ 5]  152 -400  172 -158 1000
[ 6] -228 -201 -347 -168 -404 1000
[ 7]  -37 -286 -339  -20 -485  378 1000
[ 8] -111 -437 -212  -18   79  -32 -144 1000
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: MTR.cnta
--- Number of rows: 25, columns: 8
-----

```


Total inertia: 8

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.1744E+00	+0.2718	+0.2718	02	+1.8315E+00	+0.2289	+0.5007
03	+1.2627E+00	+0.1578	+0.6586	04	+1.1800E+00	+0.1475	+0.8061
05	+7.5474E-01	+0.0943	+0.9004	06	+6.1599E-01	+0.0770	+0.9774
07	+1.8068E-01	+0.0226	+1.0000	08	+0.0000E+00	+0.0000	+1.0000

File MTR.cnvp contains the eigenvalues and relative inertia for each axis
 --- It has 8 rows and 2 columns

File MTR.cnco contains the column scores
 --- It has 8 rows and 2 columns

File :MTR.cnco

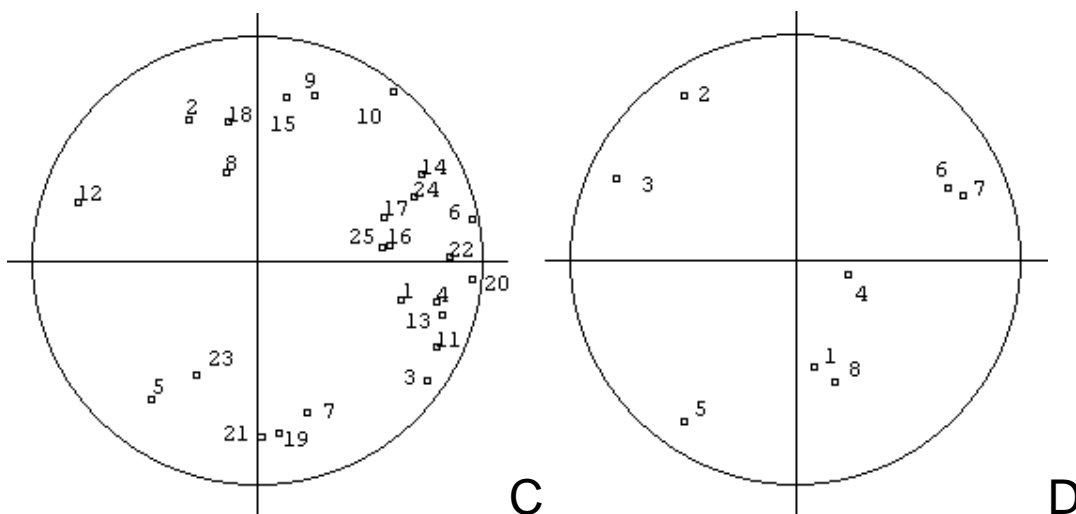
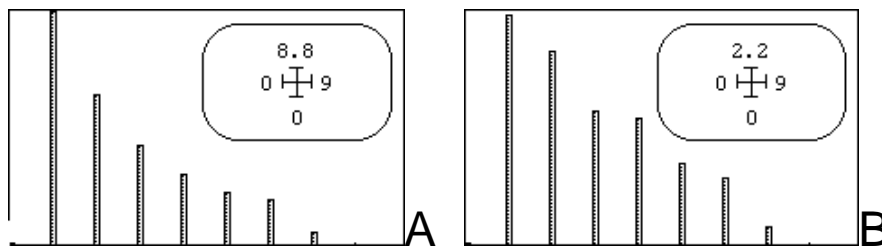
Col.	Mini	Maxi
1	-7.869e-01	7.367e-01
2	-7.147e-01	7.155e-01

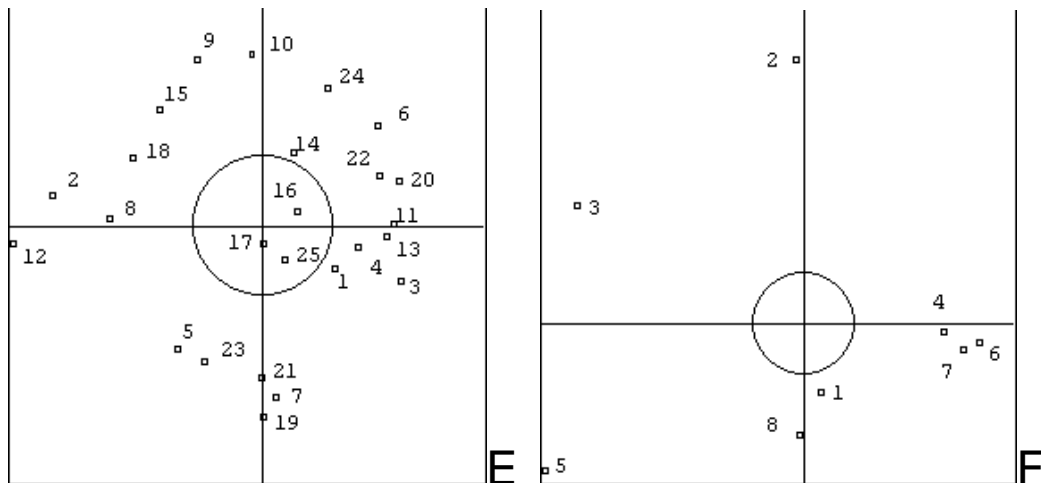
File MTR.cnli contains the row scores
 --- It has 25 rows and 2 columns

File :MTR.cnli

Col.	Mini	Maxi
1	-3.415e+00	1.874e+00
2	-2.602e+00	2.340e+00

On utilise divers modules graphiques du logiciel ADE-4 pour obtenir les figures :





- 01 Commenter les résultats du concours du point de vue des concurrents.
- 02 Laquelle des deux analyses est *a priori* contradictoire avec les conditions de l'observation ?
- 03 Sachant que les cercles ont un rayon unité, attribuer à chaque analyse les figures qui lui reviennent.
- 04 Commenter les résultats du concours au niveau de la cohérence du jury.
- 05 Que peut-on prévoir sur les résultats des versions centrées des ACP de Macon et MTR à partir des résultats obtenus par les versions normées ?
- 06 Pourquoi le programme édite dans les deux cas 8 valeurs propres alors que le nombre de variables est respectivement de 8 et 25 ?
- 07 Les jugements portés semblent ils associés à la catégorie professionnelle des juges ?
- 08 Proposer un groupement de plusieurs figures et rédiger une légende.
- 09 L'origine de la huitième valeur propre nulle est elle la même dans les deux analyses ?

6. Tableaux artificiels

DEA AMSB / Tronc commun / Décembre 1997

Un expérimentateur avisé désire se faire une opinion personnelle du comportement de l'analyse en composantes principales sur des tableaux artificiels. Il considère 3 tableaux comportant $n = 16$ lignes et $p = 16$ colonnes. Le premier est appelé **Talea** car il a été généré par une procédure de tirage aléatoire. Le second est appelé **Tgrad** car il représente une structure simple de tableau écologique définie par un gradient. Le troisième est appelé **Tparti** car il représente une structure simple de tableau écologique définie par une partition .

Talea (16-16)	Tgrad (16-16)	Tparti (16-16)
1001101000000011	1110000000000000	0101000000000000
1111101001101111	1111000000000000	1111100000000000
1111000100100110	1111100000000000	0110000100000010
1110010101000110	0011111000000000	1101000000000000
1000001000001100	0001111111000000	0110100000000000
1000001010111011	0001101111000000	0101000000000000
1011000110001011	0000111111100000	0000010111000000
0100111100000111	0000011111110000	0010011111100100
0010000111011001	0001000111000000	0000010110100000
0100110101111110	0000000111111100	0000001101100000
0010100010100001	0000000111011010	0000010101000000
0010101001010111	0000000001111110	0000000000011111
1110110010001000	0000000001011010	0000000000001101
0011010111101101	0000000000011110	0000001000010101
1111011100001100	0000000000000111	0100000001011011
0011101000101101	0000000000000111	0000000000001101

La référence écologique (16 lignes-relevés / 16 colonnes-espèces) est ici contournable, l'exercice consistant à reconnaître comment l'analyse des données repère les structures de tableaux. Si on préfère, on pourra penser que les tableaux simulent des résultats obtenus par 16 élèves (lignes) sur un test contenant 16 questions (colonnes) ou la présence-absence de 16 caractères (colonnes) sur un échantillon de 16 individus (lignes).

Les trois tableaux sont envoyés dans une simple analyse en composantes principales centrée dite encore sur matrice de covariances (PCA : Covariance matrix PCA dans ADE-4).

Exécutée trois fois la procédure donne 3 graphes de valeurs propres (Curves : Eigenvalues) :

Total inertia: 3.78906

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+7.5712E-01	+0.1998	+0.1998	02	+6.2917E-01	+0.1660	+0.3659
03	+5.2344E-01	+0.1381	+0.5040	04	+4.1266E-01	+0.1089	+0.6129
05	+3.2785E-01	+0.0865	+0.6994	06	+3.0008E-01	+0.0792	+0.7786
...							
13	+2.8836E-02	+0.0076	+0.9983	14	+5.8020E-03	+0.0015	+0.9998
15	+7.4564E-04	+0.0002	+1.0000	16	+0.0000E+00	+0.0000	+1.0000

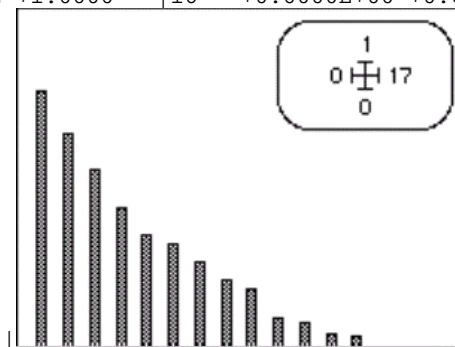


Fig 1

Total inertia: 3.26953

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.0932E+00	+0.3343	+0.3343	02	+1.0343E+00	+0.3164	+0.6507
03	+3.5646E-01	+0.1090	+0.7597	04	+2.6281E-01	+0.0804	+0.8401
05	+1.7016E-01	+0.0520	+0.8922	06	+1.3359E-01	+0.0409	+0.9330
...							
13	+8.6677E-03	+0.0027	+1.0000	14	+0.0000E+00	+0.0000	+1.0000

15 +0.0000E+00 +0.0000 +1.0000 | 16 +0.0000E+00 +0.0000 +1.0000 |

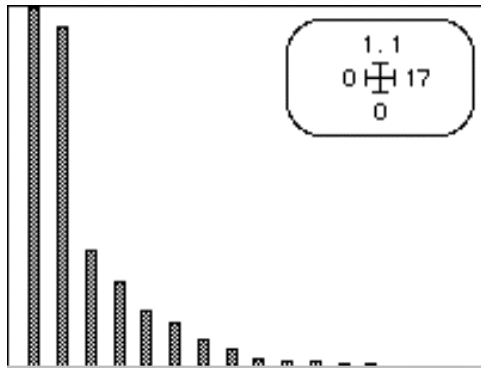


Fig 2

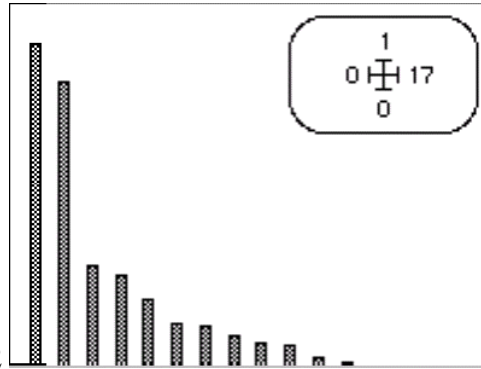


Fig 3

Total inertia: 2.85547

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+8.8946E-01	+0.3115	+0.3115	02	+7.8270E-01	+0.2741	+0.5856
03	+2.7776E-01	+0.0973	+0.6829	04	+2.4904E-01	+0.0872	+0.7701
05	+1.8449E-01	+0.0646	+0.8347	06	+1.1645E-01	+0.0408	+0.8755
...							
13	+5.4976E-03	+0.0019	+1.0000	14	+0.0000E+00	+0.0000	+1.0000
15	+0.0000E+00	+0.0000	+1.0000	16	+0.0000E+00	+0.0000	+1.0000

On obtient de même trois cartes factorielles (plan 1-2) des lignes :

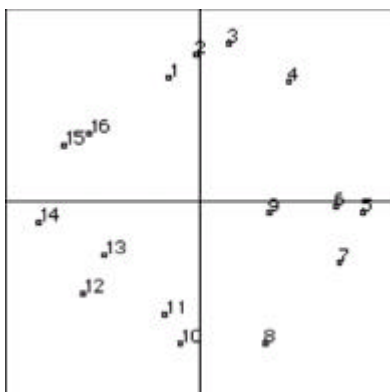


Fig 4

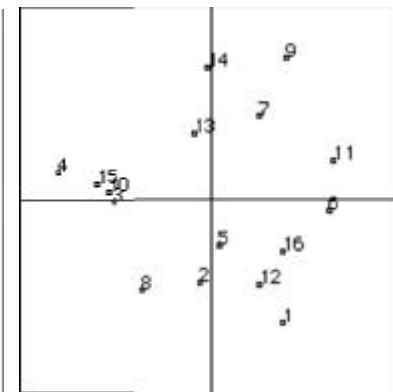


Fig 5

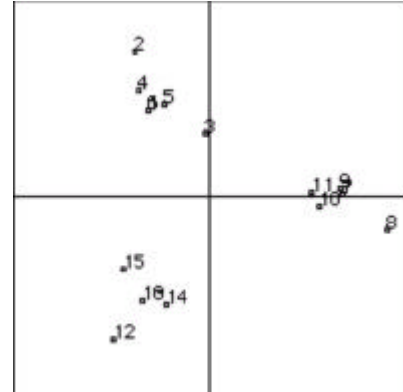


Fig 6

Les figures 4, 5 et 6 sont tracées dans une même fenêtre logique

xmin = ymin = -2, xmax = ymax = +2

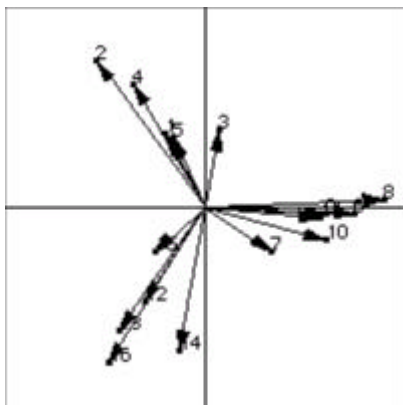


Fig 7

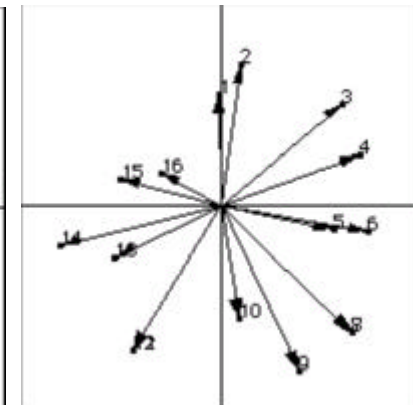


Fig 8

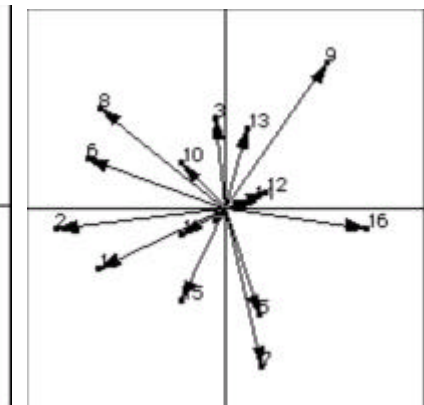


Fig 9

On obtient de même trois cartes factorielle des colonnes (plan 1-2) tracées à la même échelle

$$x_{\min} = y_{\min} = -0.5, \quad x_{\max} = y_{\max} = +0.5$$

L'expérimentateur utilise un quatrième tableau de mêmes dimensions et présentant une structure très simple. Il l'appelle Txxx et permute au hasard ses lignes et ses colonnes pour obtenir un tableau TxxxP :

TxxxP (16-16)

```
0010001001110100
0110001001110100
0110101001110010
1101001011110010
1000001011100110
1111011000100011
0011001000010010
0001000000100100
0000000000110000
0000001000110000
0001101000100000
0010111101110111
0000001000110010
0010111111111111
0011001000110010
1111111110111101
```

Le listing de l'ACP de ce tableau TxxxP donne :

```
Centered Principal Component Analysis (Pearson 1901)
Input file: TxxxP
---- Row weight:
File TxxxP.cppl contains the row weight
It has 16 rows and 1 column
Each row has 6.2500e-02 weight (Sum = 1)
---- Column weights:
File TxxxP.cppc contains the column weights
It has 16 rows and 1 column
Each column has unit weight (Sum = 16)
---- Table:
File TxxxP.cpta contains the (column) centred table
It has 16 rows and 16 columns
File :TxxxP.cpta
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-2.500e-01| 7.500e-01|
|  2|-3.125e-01| 6.875e-01|
|  3|-5.625e-01| 4.375e-01|
...
| 14|-4.375e-01| 5.625e-01|
| 15|-5.625e-01| 4.375e-01|
| 16|-2.500e-01| 7.500e-01|
|----|-----|-----|
---- Info: means and variances
File TxxxP.cpma contains the descriptive of the analysis
It contains successively:
  Number of rows: 16
  Number of columns: 16
  means and variances:
  Col.:  1 | Mean: 2.5000e-01 | Variance: 1.8750e-01
  Col.:  2 | Mean: 3.1250e-01 | Variance: 2.1484e-01
  Col.:  3 | Mean: 5.6250e-01 | Variance: 2.4609e-01
...
```

```
Col.: 14 | Mean: 4.3750e-01 | Variance: 2.4609e-01
Col.: 15 | Mean: 5.6250e-01 | Variance: 2.4609e-01
Col.: 16 | Mean: 2.5000e-01 | Variance: 1.8750e-01
```

DiagoRC: General program for two diagonal inner product analysis

Input file: TxxxP.cpta

--- Number of rows: 16, columns: 16

Total inertia: 3.02734

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+9.2776E-01	+0.3065	+0.3065	02	+4.8811E-01	+0.1612	+0.4677
03	+3.7631E-01	+0.1243	+0.5920	04	+3.5854E-01	+0.1184	+0.7104
05	+2.3706E-01	+0.0783	+0.7887	06	+1.8375E-01	+0.0607	+0.8494
...							
13	+1.6176E-02	+0.0053	+0.9996	14	+1.1927E-03	+0.0004	+1.0000
15	+0.0000E+00	+0.0000	+1.0000	16	+0.0000E+00	+0.0000	+1.0000

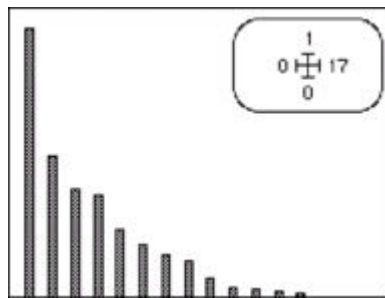


Fig 10

File TxxxP.cvpv contains the eigenvalues and relative inertia for each axis

--- It has 16 rows and 2 columns

File TxxxP.cpcv contains the column scores

--- It has 16 rows and 2 columns

File :TxxxP.cpcv

Col.	Mini	Maxi
1	-3.733e-01	4.054e-02
2	-4.214e-01	2.665e-01

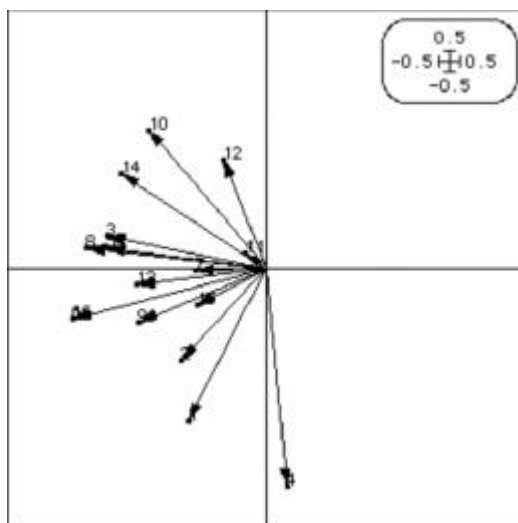


Fig 11

	1	2
1	-0.1477	-0.2254
2	-0.1625	-0.1777
3	-0.3049	0.0618
4	0.0405	-0.4214
5	-0.3066	0.0389
6	-0.3733	-0.0985
7	-0.1337	-0.0042
8	-0.3440	0.0389
9	-0.2461	-0.1043
10	-0.2275	0.2565
11	-0.0389	0.0286
12	-0.0823	0.2090
13	-0.2496	-0.0322
14	-0.2767	0.1812
15	-0.1318	-0.0717
16	-0.3733	-0.0985

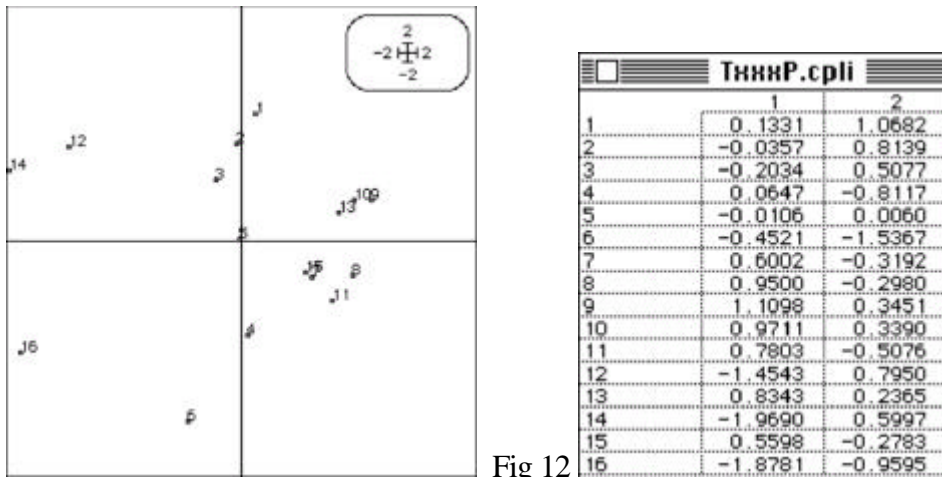


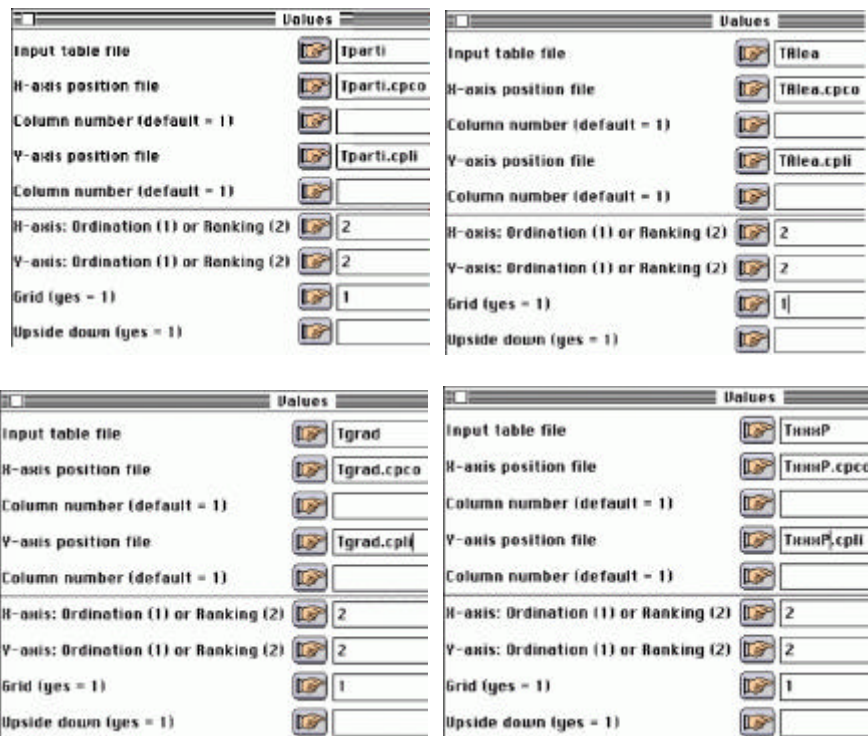
Fig 12

File TxxxP.cpli contains the row scores
 --- It has 16 rows and 2 columns
 File :TxxxP.cpli

Col.	Mini	Maxi
1	-1.969e+00	1.110e+00
2	-1.537e+00	1.068e+00

Il compare alors les résultats avec ceux obtenus par la même ACP de Txxx (non représentée).

Le module Tables : Values permet de représenter un tableau en rangeant ses lignes et ses colonnes par valeurs croissantes d'un score numérique. Utilisé quatre fois avec les paramètres :



il donne les résultats (présentés dans un ordre quelconque pour les besoins de la cause !) :

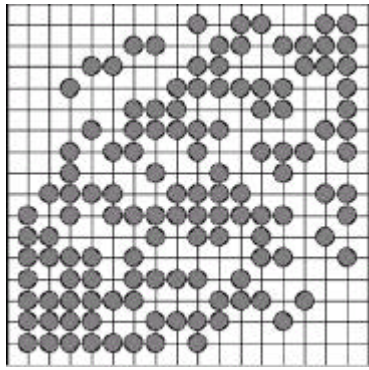


Fig 13

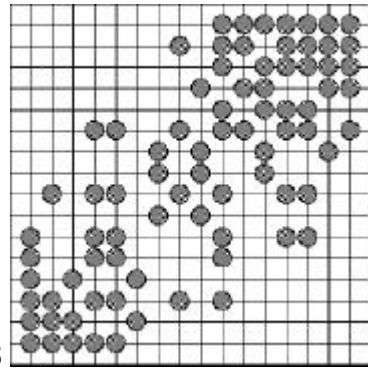


Fig 14

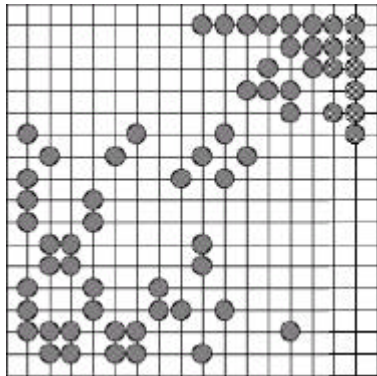


Fig 15

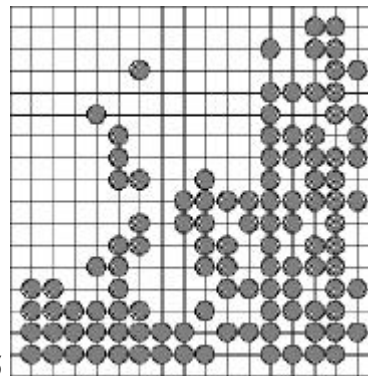


Fig 16

- 01** D'accord Pas d'accord ? La figure 1 correspond à l'analyse du tableau Talea.
- 02** D'accord Pas d'accord ? Les figures 4, 5 et 6 sont respectivement, dans l'ordre, celles des analyses de Tparti, Talea et Tgrad.
- 03** D'accord Pas d'accord ? Quand sont légitimement conservés deux axes, il existe deux faits marquants dans les données.
- 04** D'accord Pas d'accord ? Pour l'essentiel les analyses des tableaux TxxxP et Txxx sont identiques.
- 05** D'accord Pas d'accord ? Si on avait utilisés les tableaux transposés au lieu des tableaux initiaux on aurait obtenu strictement les mêmes résultats.
- 06** D'accord Pas d'accord ? Sur un tableau 16-16 on obtient toujours une valeur propre nulle.
- 07** D'accord Pas d'accord ? Il faut conserver autant d'axes que nécessaire pour obtenir au moins 50% d'inertie projetée.
- 08** D'accord Pas d'accord ? Le forme d'expression des résultats doit dépendre de la structure du tableau.
- 09** D'accord Pas d'accord ? Les données en présence-absence n'étant pas gaussiennes ces ACP sont invalides.
- 10** D'accord Pas d'accord ? Sur un vrai tableau écologique contenant un gradient simple, on risque d'avoir de sérieuses difficultés d'interprétation.

7. Europe

Soient les 12 pays :

1(a) Belgique	2(b) Danemark	3(c) Espagne
4(d) France	5(e) Grèce	6(f) Irlande
7(g) Italie	8(h) Luxembourg	9(i) Pays-Bas
10(j) Portugal	11(k) R.F. d'Allemagne	12(l) Royaume uni.

Soient les 3 secteurs de l'activité économique :

- 1(A) Agriculture, sylviculture, pêche
- 2(B) Industrie
- 3(C) Services

Chaque année la mesure des proportions de la population active civile dans les trois grands secteurs donnent un tableau à 12 lignes et 3 colonnes. On considère les tableaux de 1978 et 1986. Les données (Extrait de *Encyclopaedia Universalis*, Symposium, Les chiffres du Monde, 1989, Encyclopaedia Universalis éd., Paris, p.519.) sont :

	1978			1986		
	A	B	C	A	B	C
1*	.032	.359	.609	.028	.291	.681
2*	.079	.319	.602	.059	.282	.659
3*	.206	.372	.422	.161	.320	.519
4*	.092	.368	.540	.073	.313	.614
5*	.320	.297	.383	.285	.281	.434
6*	.206	.320	.474	.157	.287	.556
7*	.155	.381	.464	.109	.331	.560
8*	.062	.392	.546	.040	.330	.630
9*	.054	.330	.616	.049	.255	.696
10*	.313	.348	.339	.217	.348	.435
11*	.061	.444	.495	.053	.409	.538
12*	.028	.390	.582	.025	.309	.666

Question 1

Exécuter la représentation triangulaire et commenter les données.

Question 2

Notons A le tableau à 12 lignes et 3 colonnes de 1978 et B le tableau à 12 lignes et 3 colonnes de 1986. Ces deux tableaux font l'objet d'une analyse en composantes principales centrée à pondération uniforme dans R^{12} et métrique canonique dans R^3 . On note A_0 et B_0 les tableaux centrés correspondants.

Pour l'ACP de A on a :

Variable :	1	Moy :	1.3400E-01	Var :	0.1007E-01
Variable :	2	Moy :	3.6000E-01	Var :	0.1494E-02
Variable :	3	Moy :	5.0600E-01	Var :	0.7787E-02

et le message sur les valeurs propres :

/ 1:.1736E-01/.8975/0.8975/ 2:.1983E-02/.1025/1.0000/

Pour l'ACP de B on a :

Variable :	1	Moy :	1.0467E-01	Var :	0.6361E-02
Variable :	2	Moy :	3.1300E-01	Var :	0.1469E-02
Variable :	3	Moy :	5.8233E-01	Var :	0.7474E-02

et le message sur les valeurs propres :

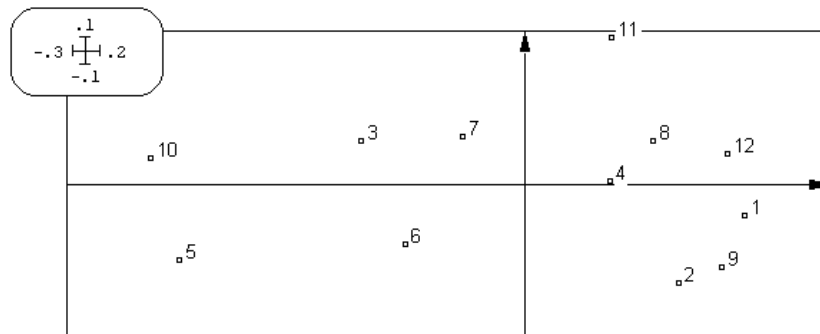
/ 1:.1318E-01/.8615/0.8615/ 2:.2119E-02/.1385/1.0000/

2.1. Placer les centres de gravité des deux nuages sur la figure 2 ci-jointe et commenter leur position respective.

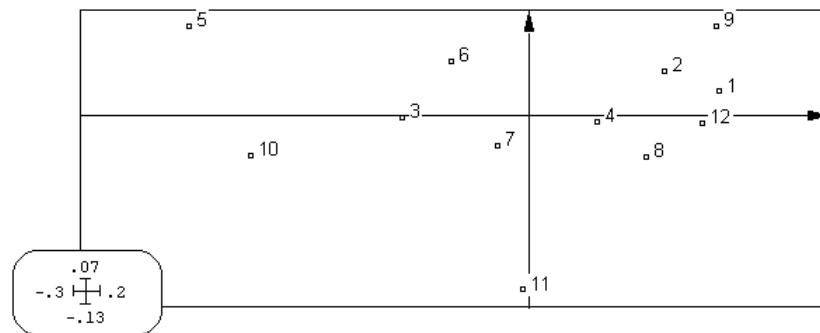
2.2. Donner l'inertie totale de chaque analyse et commenter leur valeur relative.

Question 3

La carte factorielle des lignes de l'ACP de A est :



La carte factorielle des lignes de l'ACP de B est :



3.1. Expliquer pourquoi les axes principaux des deux analyses appartiennent au plan de la représentation triangulaire.

3.2. Tracer les deux axes principaux des deux nuages de 12 points sur la figure 2 et commenter le résultat.

Question 4

4.1. Donner le rang des matrices A_0 et B_0 (justifier).

4.2. Donner les dimensions, le rang et les valeurs propres (avec leur ordre de multiplicité) des matrices $A_0 {}^t A_0$ et ${}^t A_0 A_0$, en notant ${}^t X$ la matrice transposée de la matrice X (justifier).

4.3. Chacune de ces matrices peut-elle être la matrices d'un produit scalaire (justifier) ?

4.4. Les coordonnées factorielles des colonnes des deux ACP sont respectivement :

```
A
< 1>- .9935E-01 -.1395E-01
< 2>+.1392E-01 +.3605E-01
< 3>+.8543E-01 -.2210E-01
```

```
B
< 1>- .7679E-01 +.2157E-01
< 2>-.8196E-02 -.3744E-01
< 3>+.8498E-01 +.1587E-01
```

Combien de bases du sous-espace de \mathbb{R}^3 défini par $x + y + z = 0$ peut-on former avec ces valeurs ? Parmi celles-ci combien sont-elles des bases orthogonales pour la métrique canonique ?

Ces valeurs forment deux matrices C_A et C_B à trois lignes et deux colonnes. Diagonaliser les matrices $C_A' C_A$ et $C_B' C_B$.

Sur les mêmes données

Notons A le tableau à 12 lignes et 6 colonnes obtenu en juxtaposant le tableau à 12 lignes et 3 colonnes de 1978 et le tableau à 12 lignes et 3 colonnes de 1986. Ce tableau fait l'objet d'une analyse en composantes principales centrée à pondération uniforme dans \mathbb{R}^{12} et métrique canonique dans \mathbb{R}^6 . A_0 est le tableau centré correspondant. Les résultats sont consignés dans l'annexe.

A_0 est considérée comme la matrice d'une application linéaire f de \mathbb{R}^6 dans \mathbb{R}^{12} . Donner une base orthonormée du noyau de f et son rang.

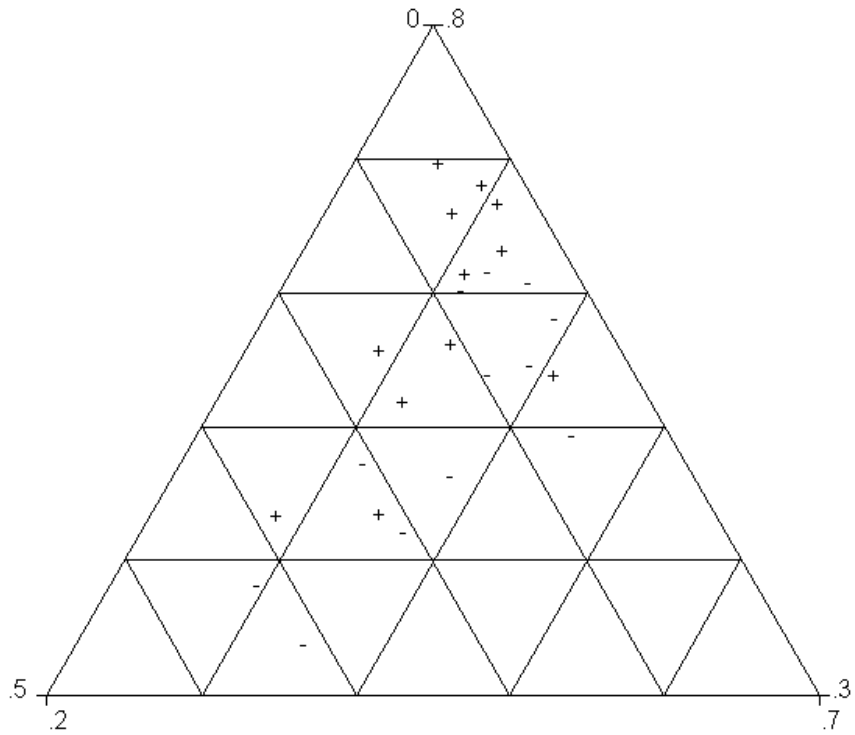


Figure 1

Placer sur la figure 1 ci-dessus ce qui y manque pour qu'elle soit utile. Rédiger une légende indiquant son origine et la nature des structures mises en évidence.

Compléter la figure 2 et rédiger une légende en indiquant les faits principaux qu'elle exprime.

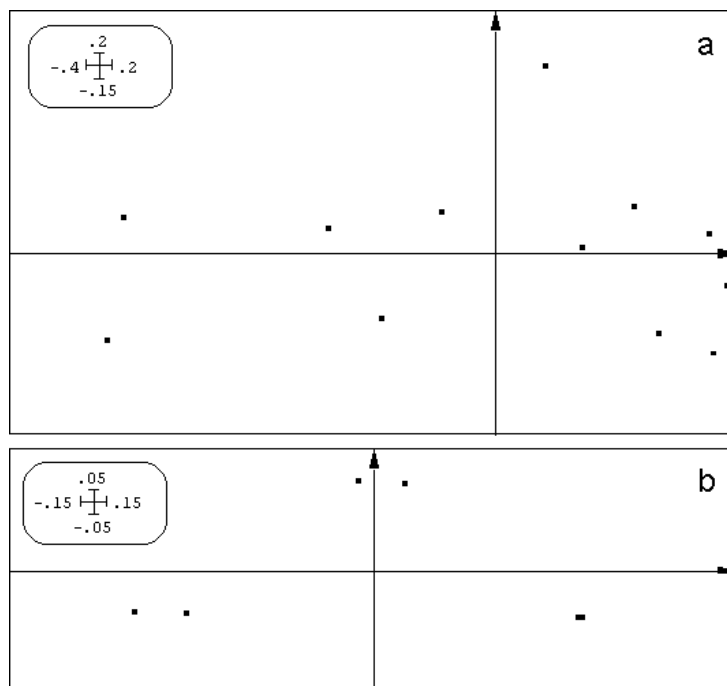


Figure 2

Annexe

Input file : Europe
Number of individuals : 12 Number of variables : 6

File Europe.cppl contains row weights
It is a 12 rows and 1 column file
Uniform weights
File Europe.cppc contains column weights which are equal to 1.
It is a 6 row and 1 column file
File Europe.cpta contains the centred table
It is a 12 row and 6 column file

File : Europe.cpta
< 1>- .1020E+00 - .1000E-02 + .1030E+00 - .7667E-01 - .2200E-01 + .9867E-01
< 2>- .5500E-01 - .4100E-01 + .9600E-01 - .4567E-01 - .3100E-01 + .7667E-01
< 3>+ .7200E-01 + .1200E-01 - .8400E-01 + .5633E-01 + .7000E-02 - .6333E-01
< 4>- .4200E-01 + .8000E-02 + .3400E-01 - .3167E-01 - .2980E-07 + .3167E-01
< 5>+ .1860E+00 - .6300E-01 - .1230E+00 + .1803E+00 - .3200E-01 - .1483E+00
< 6>+ .7200E-01 - .4000E-01 - .3200E-01 + .5233E-01 - .2600E-01 - .2633E-01
< 7>+ .2100E-01 + .2100E-01 - .4200E-01 + .4333E-02 + .1800E-01 - .2233E-01
< 8>- .7200E-01 + .3200E-01 + .4000E-01 - .6467E-01 + .1700E-01 + .4767E-01
< 9>- .8000E-01 - .3000E-01 + .1100E+00 - .5567E-01 - .5800E-01 + .1137E+00
<10>+ .1790E+00 - .1200E-01 - .1670E+00 + .1123E+00 + .3500E-01 - .1473E+00
<11>- .7300E-01 + .8400E-01 - .1100E-01 - .5167E-01 + .9600E-01 - .4433E-01
<12>- .1060E+00 + .3000E-01 + .7600E-01 - .7967E-01 - .4000E-02 + .8367E-01

Var: 1	Mean: 1.3400E-01	Var: 0.1007E-01
Var: 2	Mean: 3.6000E-01	Var: 0.1494E-02
Var: 3	Mean: 5.0600E-01	Var: 0.7787E-02
Var: 4	Mean: 1.0467E-01	Var: 0.6361E-02
Var: 5	Mean: 3.1300E-01	Var: 0.1469E-02
Var: 6	Mean: 5.8233E-01	Var: 0.7474E-02

Total inertia 3.464995E-02
File Europe.cpcl contains row contributions to the trace
It is a 12 row and 1 column file
File Europe.cpcc contains column contributions to the trace
It is a 6 row and 1 column file

----- Eigenvalues -----
/ 1: .3030E-01/.8744/0.8744/ 2: .4023E-02/.1161/0.9905
/ 3: .2231E-03/.0064/0.9970/ 4: .1054E-03/.0030/1.0000
/ 5: .0000E+00/.0000/1.0000/ 6: .0000E+00/.0000/1.0000

File Europe.cvpv contains 6 rows and 2 columns
It contains the eigenvalue and the relative inertia for each axis

File Europe.cpco contains the column scores
It is a 6 row and 2 column file
File : Europe.cpco

< 1>- .9866E-01 - .1691E-01
< 2>+ .1289E-01 + .3589E-01
< 3>+ .8577E-01 - .1898E-01
< 4>- .7736E-01 - .1775E-01
< 5>- .6408E-02 + .3730E-01
< 6>+ .8377E-01 - .1955E-01

File Europe.cpli contains the individual scores
It is a 12 row and 2 column file
File : Europe.cpli

< 1>+ .1909E+00 - .2608E-01
< 2>+ .1338E+00 - .6634E-01
< 3>- .1371E+00 + .2060E-01
< 4>+ .7046E-01 + .4654E-02

```
< 5>- .3210E+00 -.7201E-01
< 6>- .9451E-01 -.5408E-01
< 7>- .4438E-01 +.3511E-01
< 8>+.1139E+00 +.3874E-01
< 9>+.1789E+00 -.8212E-01
< 10>- .3067E+00 +.3001E-01
< 11>+.4027E-01 +.1549E+00
< 12>+.1756E+00 +.1665E-01
```

[All contributions are in 1/10000]

```
***** Absolute contributions *****
(NUM ) (FAC 1) (FAC 2) (NUM ) (FAC 1) (FAC 2)
( 1) ( 1002) ( 141) ( 1) ( 3213) ( 711)
( 2) ( 492) ( 912) ( 2) ( 55) ( 3202)
( 3) ( 517) ( 88) ( 3) ( 2428) ( 896)
( 4) ( 137) ( 4) ( 4) ( 1975) ( 783)
( 5) ( 2835) ( 1074) ( 5) ( 14) ( 3458)
( 6) ( 246) ( 606) ( 6) ( 2316) ( 950)
( 7) ( 54) ( 255)
( 8) ( 357) ( 311)
( 9) ( 880) ( 1397)
( 10) ( 2588) ( 187)
( 11) ( 45) ( 4968)
( 12) ( 848) ( 57)
```

```
***** Relatives contributions *****
(NUM ) (FAC 1) (FAC 2) (Remains) (Weight) (Cont.)
( 1) ( 9816) ( 183) ( 1) ( 833) ( 893)
( 2) ( 7833) ( 1927) ( 241) ( 833) ( 549)
( 3) ( 9579) ( 216) ( 205) ( 833) ( 472)
( 4) ( 9951) ( 43) ( 5) ( 833) ( 120)
( 5) ( 9435) ( 475) ( 90) ( 833) ( 2627)
( 6) ( 7496) ( 2454) ( 50) ( 833) ( 287)
( 7) ( 5647) ( 3534) ( 818) ( 833) ( 84)
( 8) ( 8922) ( 1031) ( 46) ( 833) ( 350)
( 9) ( 8252) ( 1739) ( 9) ( 833) ( 933)
( 10) ( 9840) ( 94) ( 66) ( 833) ( 2300)
( 11) ( 615) ( 9100) ( 285) ( 833) ( 634)
( 12) ( 9856) ( 89) ( 55) ( 833) ( 752)
```

```
(NUM ) (FAC 1) (FAC 2) (Remains) (Weight) (Cont.)
( 1) ( 9670) ( 284) ( 45) (10000) ( 2905)
( 2) ( 1113) ( 8625) ( 262) (10000) ( 431)
( 3) ( 9447) ( 463) ( 90) (10000) ( 2247)
( 4) ( 9407) ( 495) ( 97) (10000) ( 1836)
( 5) ( 280) ( 9471) ( 250) (10000) ( 424)
( 6) ( 9388) ( 511) ( 101) (10000) ( 2157)
```