

Fiche de Biostatistique

Problèmes d'analyse des données

D. Chessel

Résumé

La fiche contient 5 problèmes d'analyse des données.

Plan

1.	DIFFERENTIEATEURS SEMANTIQUES.....	2
2.	NOTES D'EXAMEN.....	7
3.	VEUVAGE.....	14
4.	TRUITES.....	22
5.	KYSTE DE L'OVAIRE	28

1. Différentiateurs sémantiques

DEA AMSB / Juin 1991

On a demandé à 11 étudiants ce qu'ils pensaient de 15 disciplines scientifiques au moyen de 6 paires d'adjectifs antonymes. Les 11 étudiants appartiennent au DEA de didactique des disciplines scientifiques et sont ou seront des enseignants scientifiques. Ils sont simplement connus par leur numéro d'ordre (1 à 11) car **on ne s'intéresse ici qu'à leur opinion collective.**

Les 15 disciplines sont :

1-algèbre	6-éthologie	b-physique nucléaire
2-astrologie	7-informatique	c-psychologie
3-biologie moléculaire	8-linguistique	d-science
4-didactique	9-médecine	e-sociologie
5-écologie	a-neurologie	f-technologie

Les 6 paires d'adjectifs utilisés sont appelées des différentiateurs sémantiques d'OSGOOD. Ce sont :

- I précis (1)-imprécis (5)
- II dur (1)-mou (5)
- III subjectif (1)-objectif (5)
- IV faux (1)-vrai (5)
- V faible (1)-fort (5)
- VI fantaisiste (1)-sérieux (5)

Pour un étudiant donné, une discipline donnée et une paire d'adjectifs donnée l'opinion exprimée sur la discipline par l'étudiant au moyen du différentiateur est une note qui peut prendre 5 valeurs :

- 1— association forte avec le premier terme du différentiateur
- 2— préférence pour le premier terme du différentiateur
- 3— absence d'opinion
- 4— préférence pour le second terme du différentiateur
- 5— association forte avec le second terme du différentiateur

Pour chacun des 6 adjectifs on a donc un tableau des notes choisies (15 disciplines-lignes et 11 étudiants/colonnes).

Les données communiquées par V. Ndiaye et P. Clément sont dans l'annexe 1. Dans ce problème, aucune question n'est obligatoire en ce sens qu'on n'est pas obligé d'exprimer une opinion sur tous les points de vue.

Question 1

On choisit de calculer la note moyenne utilisée par les 11 personnes interrogées pour chaque couple discipline-différentiateur. On obtient un tableau 15-11 dont on trouvera le contenu et l'ACP dans l'annexe 2.

1.1 — Peux-t'on savoir en utilisant les seuls résultats de l'annexe de quel type d'ACP on s'est servi ?

1.2 — Indiquer comment vous obtiendriez ces résultats en partant du fichier ASCII des 6 tableaux de l'annexe 1 superposés (90 lignes et 11 caractères par lignes).

1.3 — Cette analyse contribue-t'elle à la lecture des résultats numériques. Si non, pourquoi ? Si oui, commenter l'information obtenue ?

1.4 — Obtiendrait-on le même résultat en faisant l'ACP du tableau transposé à 11 lignes et 15 colonnes ? Si non, le résultat serait il meilleur ou moins bon, du point de vue de l'objectif poursuivi ?

Question 2

On choisit de compter, pour chaque couple discipline-différentiateur, le nombre de personnes interrogées ayant utilisé respectivement les notes 1,2, ..., 5. On obtient un tableau 15-30 dont on trouvera le contenu et l'AFC dans l'annexe 3.

2.1 — Indiquer comment vous obtiendriez ces résultats en partant du même fichier.

2.2 — Les coordonnées factorielles des colonnes de cette analyse ont-elles des propriétés particulières ?

2.3 — Que gagnerait-on à utiliser une analyse des correspondances intra-classes en indiquant la présence de 6 groupes de 5 colonnes ?

2.4 — Cette analyse contribue-t'elle à la lecture des résultats numériques. Si non, pourquoi ? Si oui, commenter l'information obtenue ?

Question 3

3.1 — Les deux analyses précédentes utilisent-elles les mêmes pondérations des individus (justifier) ?

3.2 — Les deux analyses précédentes peuvent-elles être considérées comme des parties d'analyses inter-classes (justifier) ?

3.3 — Les deux analyses vous paraissent-t'elles convergentes, divergentes, contradictoires ? Les deux analyses vous paraissent-t'elles d'égal intérêt pour l'objectif poursuivi ?

3.4 — Proposer une ou plusieurs manières de voir ou de mesurer l'adéquation des opinions personnelles à l'opinion collective.

Annexe 1 : données brutes

précis	dur	subjectif	faux	faible	fantaisiste	
11111112111	12322354342	54355454554	54355455551	54355455551	55352454331	55254454545
34552554523	55352453523	11111111132	31212112144	32131112144	32131112144	31112114132
31111222123	13313213111	43452444532	44454444553	41553324334	45554412553	45554412553
34234454425	44325353335	45241214131	34443414543	15333312551	55542434551	55542434551
32225443242	32332443333	43541243435	53553244554	34343114334	54422243525	54422243525
43224422333	13331342335	53241244311	33453444333	33333344333	33454242123	33454242123
13115112111	13322212221	54454424453	54445433353	54345444555	54454444555	54454444555
32213442234	34332314134	43451344411	44454443533	44344344332	44453342532	44453342532
33123444514	51324312511	31442221152	22443224352	33342322354	33554122452	33554122452
23112222432	13322312332	43553244442	53444443433	43342323333	53554442534	53554442534
33121222213	21321212322	33354424552	33344444542	55344454334	34554454554	34554454554
43444454443	53343354534	32311112221	33442224443	33334323341	33442224142	33442224142
12212222112	22232312131	44444444552	45445444553	44444344354	55554442552	55554442552
34324454542	43342314334	43242213222	33343323433	34342223332	33342322532	33342322532
13212211112	12222253131	53553254554	53444344553	55334344354	54544443535	54544443535

Annexe 2 : Analyse en composantes principales

```
Lig 1/1.091/2.818/4.455/4.273/3.636/4.364/
Lig 2/3.909/3.818/1.273/2.000/2.091/1.818/
Lig 3/1.727/1.818/3.636/4.182/3.364/3.909/
Lig 4/3.636/3.636/2.545/3.545/2.909/3.909/
Lig 5/3.000/3.000/3.455/4.091/3.000/3.455/
Lig 6/2.909/2.818/2.727/3.545/3.182/3.000/
Lig 7/1.636/1.909/4.000/3.909/4.364/4.455/
Lig 8/2.727/2.818/3.091/3.909/3.455/3.545/
Lig 9/3.091/2.545/2.455/3.000/3.091/3.273/
Lig 10/2.182/2.273/3.636/3.727/3.000/4.000/
Lig 11/2.000/1.909/3.636/3.636/4.000/4.364/
Lig 12/3.909/3.818/1.727/3.182/2.909/2.818/
Lig 13/1.636/2.000/4.000/4.273/3.909/4.182/
Lig 14/3.636/3.091/2.455/3.091/2.818/2.909/
Lig 15/1.545/2.182/4.182/4.000/3.909/...../
```

Number of individuals : 15 Number of variables : 6

Means and variances :

```
Var: 1 Mean: 2.5758E+00 Var:0.8415E+00
Var: 2 Mean: 2.6970E+00 Var:0.4437E+00
Var: 3 Mean: 3.1515E+00 Var:0.8051E+00
Var: 4 Mean: 3.6242E+00 Var:0.3470E+00
Var: 5 Mean: 3.3091E+00 Var:0.3160E+00
Var: 6 Mean: 3.6121E+00 Var:0.5118E+00
```

Total inertia 3.265161

```
----- Eigenvalues -----
/ 1:...../...../ 2:.1820E+00/.0557/0.9249
/ 3:.1130E+00/.0346/0.9595/ 4:.6725E-01/.0206/0.9801
/ 5:.4695E-01/.0144/0.9945/ 6:.1809E-01/.0055/1.0000
-----
```

```
File : C.cpli File : C.cpc
< 1>-.2010E+01 -.5213E+00 < 1>+.8832E+00 -.1181E+00
< 2>+.3626E+01 +..... < 2>+.5719E+00 -.2932E+00
< 3>-.1300E+01 +.3414E+00 < 3>-.8802E+00 -.9519E-01
< 4>+.1217E+01 -.8970E+00 < 4>-.5132E+00 -.2285E+00
< 5>+.1787E+00 -.6227E+00 < 5>-.4994E+00 +.4173E-01
< 6>+.7412E+00 +.1466E+00 < 6>-.6658E+00 -.1378E+00
< 7>-.1936E+01 +.2907E+00
< 8>+.4871E-01 -.2286E+00
< 9>+.9718E+00 +.5398E+00
< 10>-.6970E+00 +.8172E-01
< 11>-.1328E+01 +.4111E+00
```

< 12>+.2391E+01 -.3676E+00
 < 13>-.1773E+01 +.7701E-01
 < 14>+.1640E+01 +.5599E-01
 < 15>-.1771E+01 +.8274E-01

Annexe 3 : Analyse des correspondances

1/•/1/ / / /1/4/3/2/1/ / /1/4/6/1/ /1/2/7/1/1/3/2/4/ /1/ /4/6/
 2/ /2/2/2/5/ /2/3/1/5/9/1/1/ / /5/3/1/2/ /5/2/2/2/ /6/2/2/1/ /
 3/5/4/2/ / /6/1/4/ / / /2/2/5/2/ / /1/7/3/1/1/4/3/2/1/1/1/3/5/
 4/ /2/2/5/2/ /1/5/2/3/4/2/1/3/1/1/ /3/6/1/3/1/4/ /3/1/1/1/3/5/
 5/ /5/2/3/1/ /2/7/2/ /1/1/3/4/2/ /1/2/3/5/2/ /5/4/ / /4/1/3/3/
 6/ /4/4/3/ /2/1/6/1/1/3/2/2/3/1/ / /6/4/1/ / /9/2/ /1/3/3/3/1/
 7/8/1/1/ /1/3/6/2/ / / /1/1/6/3/ / /4/4/3/ / /1/5/5/ / / /6/5/
 8/1/4/3/3/ /2/1/5/3/ /3/ /2/5/1/ / /3/6/2/ /1/4/6/ / /2/3/4/2/
 9/2/1/3/4/1/4/2/2/1/2/3/4/1/2/1/ /5/2/3/1/ /3/5/2/1/1/3/2/2/3/
 a/2/6/2/1/ /2/4/5/ / / /2/2/5/2/ / /4/6/1/ /2/7/2/ / /1/2/4/4/
 b/3/5/3/ / /3/6/2/ / / /2/3/3/3/ /1/3/6/1/ / /3/5/3/ / /1/5/5/
 c/ / /2/8/1/ / /5/3/3/5/4/2/ / / /3/3/5/ /1/1/7/2/ /1/4/2/4/ /
 d/4/7/ / / /3/5/3/ / / /1/ /8/2/ / /1/6/4/ / /2/8/1/ /2/ /3/6/
 e/ /2/2/5/2/1/1/5/4/ /1/6/2/2/ / /1/8/2/ / /4/5/2/ / /4/5/1/1/
 f/6/4/1/ / /3/5/2/ /1/ /1/2/2/6/ / /3/5/3/ / /4/4/3/ / /2/5/4/

Total number : 990

----- Eigenvalues -----
 / 1: .3271E+00/.4455/0.4455/ 2: .1567E+00/.2134/0.6589
 / 3: .5816E-01/.0792/0.7381/ 4: .4632E-01/.0631/0.8011
 / 5: .3158E-01/.0430/0.8442/ 6: .3022E-01/.0411/0.8853
 / 7: .2120E-01/.0289/0.9142/ 8: .1975E-01/.0269/0.9411
 / 9: .1505E-01/.0205/0.9616/ 10: .9069E-02/.0124/0.9739
 / 11: .8185E-02/.0111/0.9851/ 12: .5534E-02/.0075/0.9926
 / 13: .3736E-02/.0051/0.9977/ 14: .1706E-02/.0023/1.0000
 / 15: .1193E-07/.0000/1.0000

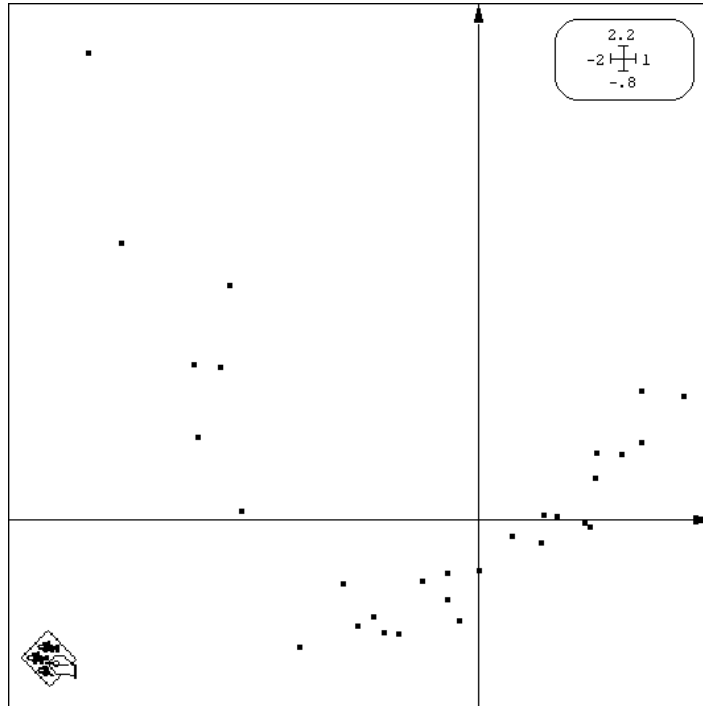
Total inertia .7343216
 File : D.fcli

 < 1>+.6057E+00 +.5083E+00
 < 2>-.1370E+01 +.9743E+00
 < 3>+.3905E+00 +.7450E-01
 < 4>-.4161E+00 +.1398E+00
 < 5>-.2251E-01 -.1777E+00
 < 6>-.2079E+00 -.4080E+00
 < 7>+.6824E+00 +.3208E+00
 < 8>+.1013E-01 -.3041E+00
 < 9>-.3490E+00 -.1551E+00
 < 10>+.2400E+00 -.2503E+00
 < 11>+.4587E+00 +.3148E-01
 < 12>-.7431E+00 -.3669E+00
 < 13>+.6071E+00 +.1103E+00
 < 14>-.4476E+00 -.6766E+00
 < 15>+.5620E+00 +.1793E+00

 File : D.fcco

 < 1>+.8716E+00 +.5248E+00 < 16>-.1664E+01 +.1992E+01
 < 2>+.2681E+00 -.9183E-01 < 17>-.1011E+01 +.4040E-01
 < 3>-.2391E+00 -.2597E+00 < 18>-.7901E-01 -.4232E+00
 < 4>-.7622E+00 -.5358E+00 < 19>+.1458E+00 -.6744E-01
 < 5>-.1212E+01 +.6644E+00 < 20>+.6123E+00 +.2829E+00
 < 6>+.4733E+00 -.2870E-01 < 21>-.1061E+01 +.1001E+01
 < 7>+.5083E+00 +.2897E+00 < 22>-.5738E+00 -.2640E+00
 < 8>-.1301E+00 -.2206E+00 < 23>-.1332E+00 -.3332E+00
 < 9>-.5137E+00 -.4478E+00 < 24>+.3357E+00 +.1548E-01

< 10>-.1100E+01 +.6567E+00	< 25>+.6905E+00 +.5503E+00
< 11>-.1195E+01 +.3517E+00	< 26>-.1518E+01 +.1178E+01
< 12>-.3415E+00 -.4848E+00	< 27>-.4502E+00 -.4097E+00
< 13>+.4857E-02 -.2178E+00	< 28>-.4015E+00 -.4738E+00
< 14>+.4547E+00 -.1384E-01	< 29>+.2807E+00 +.2628E-01
< 15>+.6934E+00 +.3342E+00	< 30>+.5005E+00 +.1803E+00



2. Notes d'examen

Deug MASS 2° année / Mai 96

Dans les temps anciens, les étudiants de la filière D2F5 passaient un examen de fin d'années (session de Juin) et le jury regroupait les notes obtenues pour $p = 9$ matières, à savoir :

- 1- Algèbre et Analyse des données (sur 100)
- 2- Analyse (sur 60)
- 3- Probabilités (sur 80)
- 4- Informatique (sur 60)
- 5- Dominante (Sociologie ou Économie sur 120)
- 6- Options (sur 40)
- 7- Ouvertures (sur 40)
- 8- Anglais (sur 40)
- 9- Éducation physique et sportive (bonification ≤ 15)

A l'issue de la délibération, il y avait 6 possibilités :

- A- Éliminé
- B- Éliminé après l'oral de rattrapage
- C- Admis après l'oral de rattrapage
- D- Admis avec la mention Passable
- E- Admis avec mention AB
- F- Admis avec mention B

On considère un ensemble de $n = 104$ étudiants. Les notes obtenues sont normalisées, c'est à dire subissent la transformation :

$$y_{ij} = \frac{x_{ij} - \frac{1}{n} \sum_{i=1}^n x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(x_{ij} - \frac{1}{n} \sum_{i=1}^n x_{ij} \right)^2}}$$

Ces données transformées sont éditées dans les tableaux 1 et 2. Tous les nuages de points bivariés sont représentés dans la figure 1. On associe au tableau des données transformées une pondération uniforme des lignes et une pondération unitaire des colonnes. Le triplet obtenu donne un Analyse en Composantes Principales (ACP) normée.

On obtient successivement :

— la matrice de corrélation :

```
-----
[ 1] 1000
[ 2] 445 1000
[ 3] 504 516 1000
[ 4] 389 320 373 1000
[ 5] 366 207 167 76 1000
-----
```

```
[ 6] 537 404 444 250 371 1000
[ 7] 196 62 112 86 339 204 1000
[ 8] 114 -120 187 131 406 94 21 1000
[ 9] 235 158 269 62 178 259 77 137 1000
```

— les valeurs propres :

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+3.1014E+00	+0.3446	+0.3446	02	+1.3630E+00	+0.1514	+0.4960
03	+1.0323E+00	+0.1147	+0.6107	04	+9.3405E-01	+0.1038	+0.7145
05	+7.3975E-01	+0.0822	+0.7967	06	+5.7467E-01	+0.0639	+0.8606
07	+5.3254E-01	+0.0592	+0.9197	08	+4.3754E-01	+0.0486	+0.9684
09	+2.8478E-01	+0.0316	+1.0000				

— les coordonnées des colonnes :

File Res.cnco contains the column scores
 --- It has 9 rows and 2 columns
 File :Res.cnco

Col.	Mini	Maxi
1	2.755e-01	7.925e-01
2	-6.698e-01	4.626e-01

— les coordonnées des lignes :

File Res.cnli contains the row scores
 --- It has 104 rows and 2 columns
 File :Res.cnli

Col.	Mini	Maxi
1	-5.727e+00	4.265e+00
2	-3.805e+00	3.617e+00

— les contributions absolues des variables :

Num	Fac 1	Fac 2
1	2024	62
2	1375	1570
3	1770	430
4	901	580
5	989	2882
6	1773	1
7	358	1037
8	244	3291
9	561	143

— les contributions relatives des variables :

Num	Fac 1	Fac 2	Remains	Weight	Cont.
1	6280	85	3634	10000	1111
2	4266	2140	3593	10000	1111
3	5491	586	3922	10000	1111
4	2795	791	6413	10000	1111
5	3067	3928	3003	10000	1111
6	5499	2	4497	10000	1111
7	1112	1413	7473	10000	1111
8	759	4486	4754	10000	1111
9	1740	195	8064	10000	1111

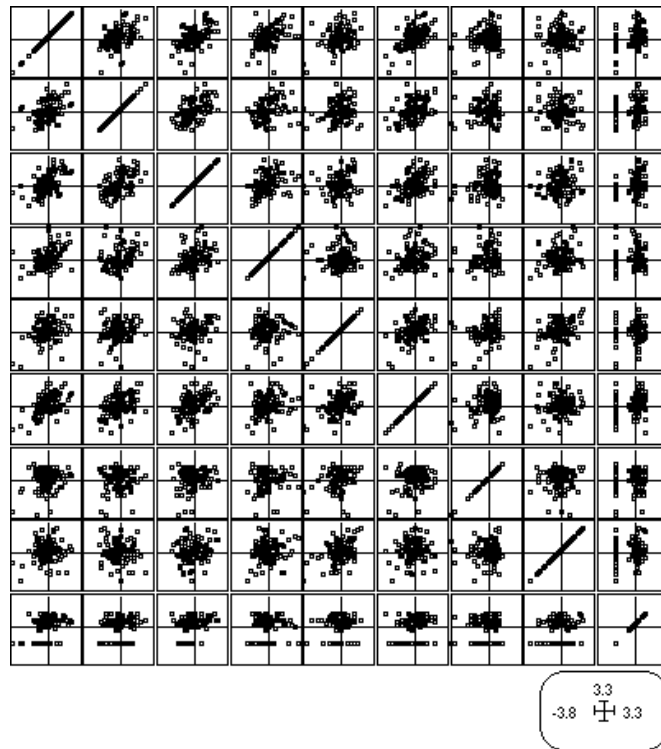


Figure 1

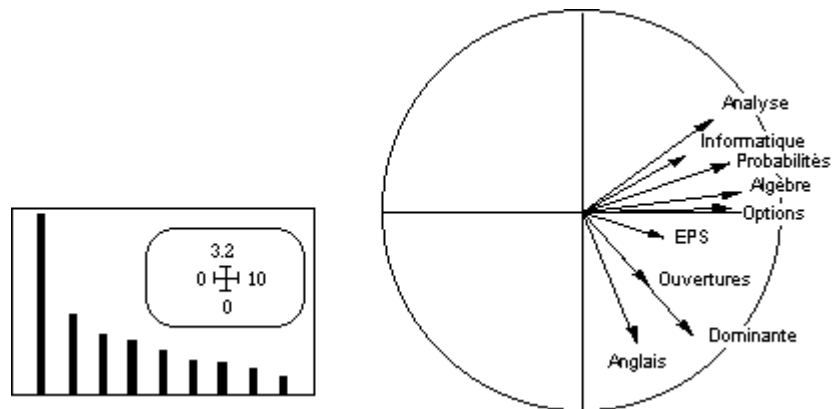


Figure 2

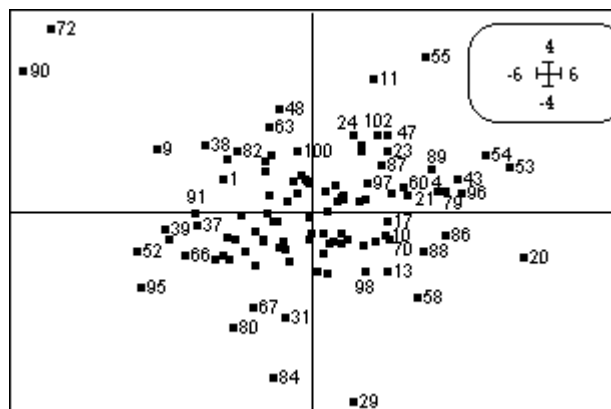


Figure 3

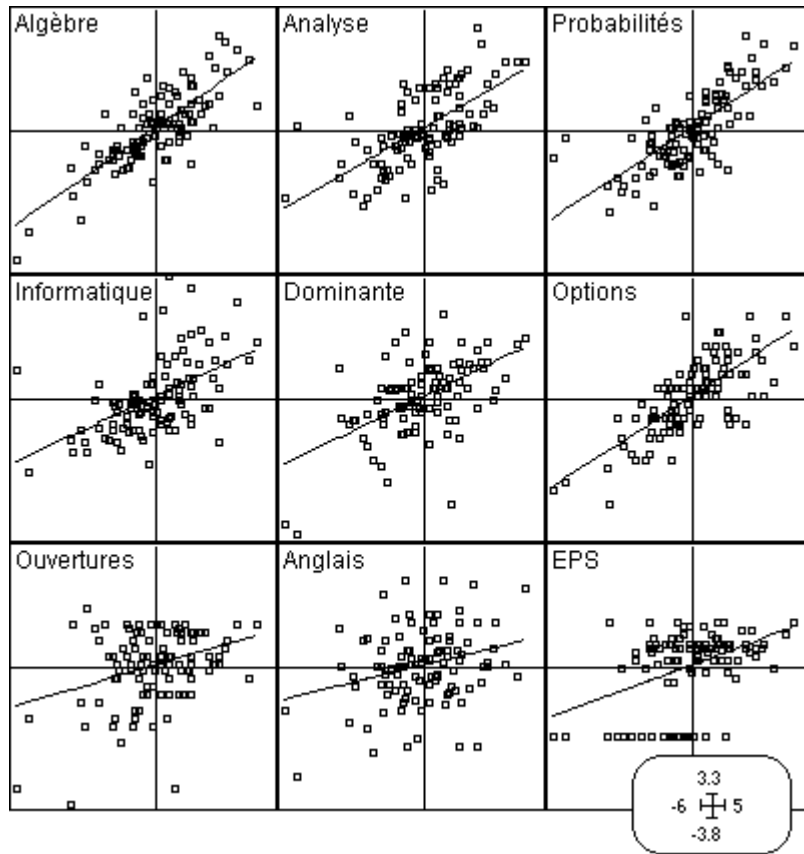


Figure 4

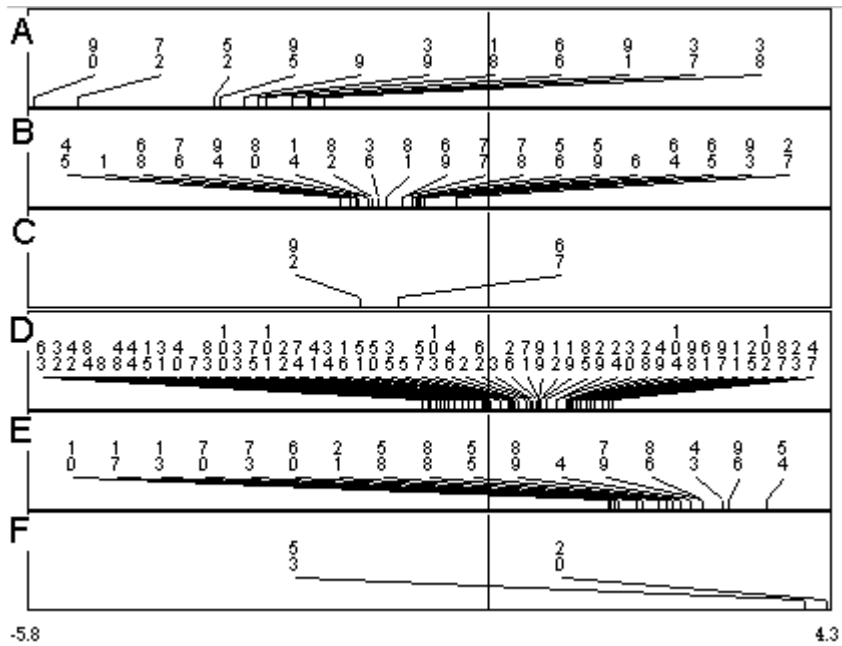


Figure 5

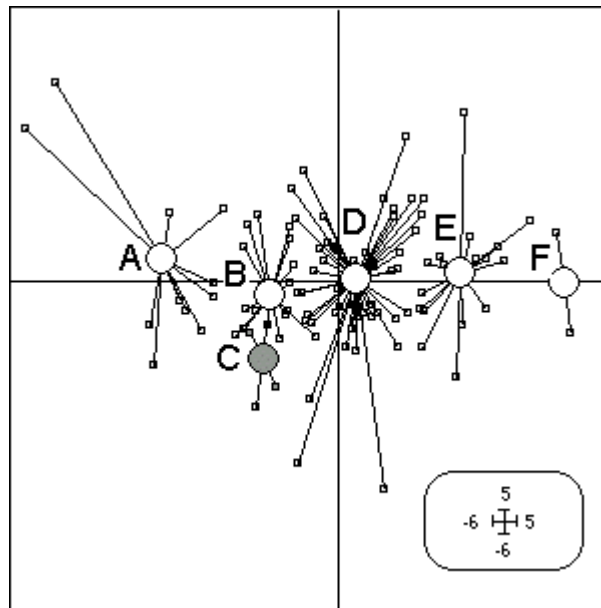


Figure 6

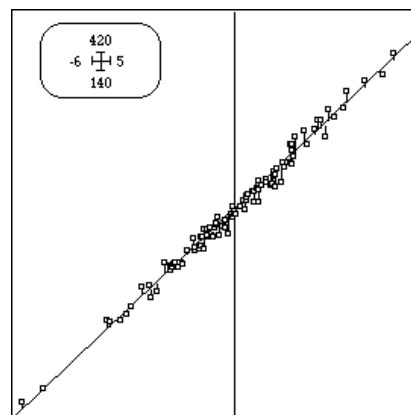


Figure 7

01 D'accord Pas d'accord ? La normalisation des données élimine l'effet des échelles de notation.

02 Quels faits sont-ils mis en évidence dans la figure 1 ?

03 D'accord Pas d'accord ? Connaissant le note obtenue en dominante, on ne peut jamais expliquer plus de 20% de la variance de la note obtenue dans une autre matière par une régression linéaire.

04 D'accord Pas d'accord ? La matrice de corrélation est de rang 8.

05 D'accord Pas d'accord ? Seul le premier axe principal doit faire l'objet d'une interprétation.

06 D'accord Pas d'accord ? La cohésion du jury est très forte.

07 D'accord Pas d'accord ? Numériquement, la matière Algèbre et Analyse des Données est la plus représentative de la valeur globale des étudiants.

08 D'accord Pas d'accord ? La valeur des coefficients joue un rôle limité dans la mesure de la valeur des étudiants.

09 D'accord Pas d'accord ? Globalement 35% de la variabilité des notes sert à l'élaboration du résultat final et 65% de cette variabilité est l'expression de la diversité entre les personnalités des étudiants.

10 Donner une légende mathématique pour la partie droite de la figure 2.

3. Veuvage

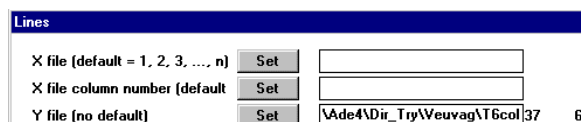
DEA AMSB / Tronc commun / Décembre 1996

L'INSEE propose un tableau donnant le pourcentage d'hommes veufs en fonction de l'âge et de la catégorie socio-professionnelle. En colonnes figurent les catégorie socio-professionnelle (1- Agriculteurs, 2- Artisans, 3- Cadres supérieurs, 4- Professions intermédiaires, 5- Employés et 6- Ouvriers). En lignes, on trouve l'âge des personnes à partir de 50 ans :

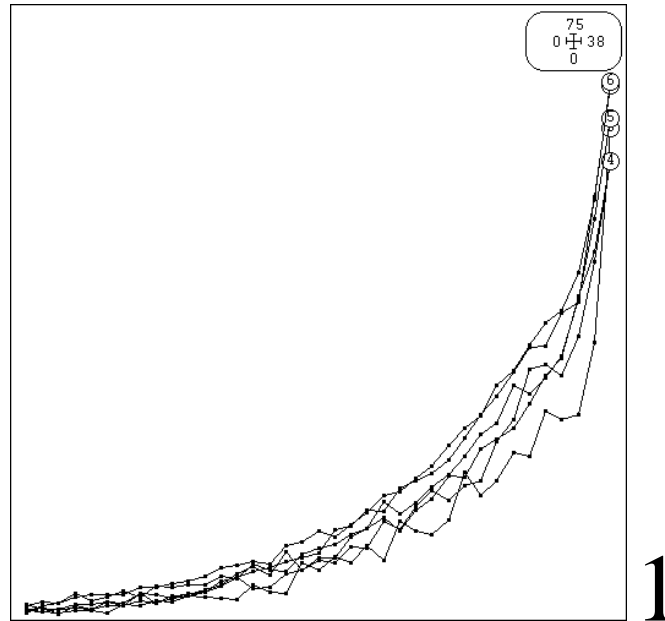
Age	1- Agriculteurs	2- Artisans	3- Cadres Sup	4- Cadres Moy	5- Employés	6- Ouvriers
50	2.06	1.00	1.25	1.37	1.13	1.83
51	1.42	1.72	1.19	1.15	1.85	2.25
52	1.34	.82	1.04	1.41	2.11	2.23
53	2.05	1.92	1.22	1.63	3.32	2.97
54	2.08	1.43	1.30	1.40	2.51	3.16
55	2.29	2.37	.97	1.84	2.97	3.20
56	1.80	2.19	2.07	1.88	3.59	3.18
57	3.35	3.32	1.88	2.39	2.89	4.10
58	3.14	2.09	2.43	2.21	4.26	4.05
59	2.71	3.11	2.31	2.79	4.16	4.51
60	3.42	3.16	3.06	3.04	4.37	4.83
61	3.47	3.86	2.92	3.43	4.33	5.45
62	4.30	4.39	2.72	4.82	5.42	6.52
63	5.27	5.67	2.58	5.23	5.08	6.70
64	5.97	6.62	4.38	3.95	6.86	7.21
65	6.51	5.63	3.52	4.08	6.18	6.88
66	7.21	8.37	3.38	5.67	6.02	9.20
67	7.63	6.15	7.08	6.14	8.15	9.62
68	8.25	7.63	6.15	7.35	8.82	10.89
69	11.12	7.72	7.77	7.14	9.36	10.24
70	11.54	10.16	7.14	9.06	10.51	11.71
71	13.45	11.27	9.20	8.87	11.27	13.15
72	13.38	12.56	7.33	12.20	14.50	15.20
73	16.19	11.00	12.15	11.03	13.07	15.77
74	17.01	13.45	10.90	13.86	14.40	17.33
75	17.86	14.78	10.42	15.67	16.34	18.60
76	19.55	17.70	12.26	14.70	17.85	21.33
77	22.19	17.49	18.04	16.43	19.97	23.46
78	25.12	20.91	15.31	17.03	22.66	24.91
79	27.28	22.08	16.99	21.80	24.03	28.70
80	30.34	23.44	20.44	24.42	28.71	30.42
81	33.33	26.46	20.00	30.55	27.55	33.63
82	33.41	29.79	25.54	31.18	29.52	36.20
83	37.40	31.65	24.45	29.80	32.28	37.77
84	38.82	39.57	25.00	34.67	39.22	42.38
[85,89]	51.60	44.94	33.91	43.64	48.89	51.29
>=90	65.03	55.87	59.89	55.95	61.12	65.59

Ce tableau donne deux fichiers. Le premier T6col a 37 lignes et 6 colonnes. Le second T6lig a 6 lignes et 37 colonnes est obtenu par transposition.

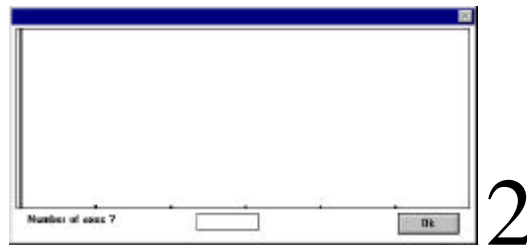
Le module Curves : Lines :



donne la figure 1 :



Le tableau T6col est envoyé dans un programme d'ACP centrée par colonnes (PCA: Covariance matrix PCA) :



```
Centered Principal Component Analysis (Pearson 1901)
Input file: T6col
---- Row weights:
File T6col.copl contains the row weights. It has 37 rows and 1 column. Each row
has 2.7027e-02 weight (Sum = 1)
---- Column weights:
File T6col.copc contains the column weights. It has 37 rows and 1 column. Each
column has unit weight (Sum = 6)
---- Table:
File T6col.cpta contains the (column) centred table
It has 37 rows and 6 columns
```

```
---- Info: means and variances
File T6col.cpma contains the descriptive of the analysis
It contains successively:
Number of rows: 37
Number of columns: 6
means and variances:
Col.: 1 | Mean: 1.5105e+01 | Variance: 2.3296e+02
Col.: 2 | Mean: 1.3040e+01 | Variance: 1.7411e+02
Col.: 3 | Mean: 1.0492e+01 | Variance: 1.3974e+02
Col.: 4 | Mean: 1.2432e+01 | Variance: 1.7158e+02
Col.: 5 | Mean: 1.4196e+01 | Variance: 1.9306e+02
Col.: 6 | Mean: 1.6126e+01 | Variance: 2.3081e+02
```

DiagoRC: General program for two diagonal inner product analysis

Input file: T6col.cpta
 --- Number of rows: 37, columns: 6

Total inertia: 1142.27

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.1326E+03	+0.9915	+0.9915	02	+6.1744E+00	+0.0054	+0.9969
03	+1.4015E+00	+0.0012	+0.9982	04	+1.1847E+00	+0.0010	+0.9992
05	+6.5321E-01	+0.0006	+0.9998	06	+2.6492E-01	+0.0002	+1.0000

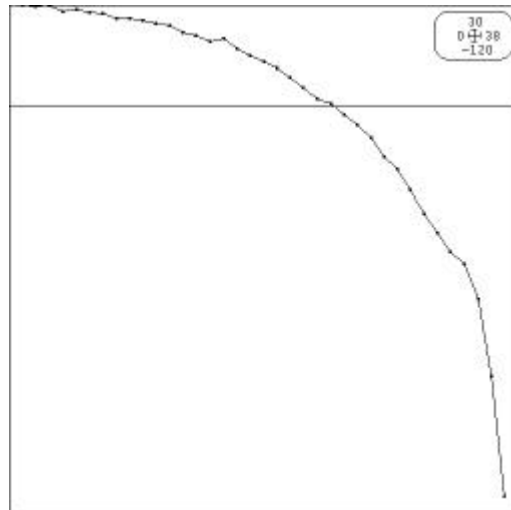
File T6col.cvpv contains the eigenvalues and relative inertia for each axis
 --- It has 6 rows and 2 columns

File T6col.cpcv contains the column scores
 --- It has 6 rows and 1 columns

Col.	Mini	Maxi
1	-1.524e+01	-1.160e+01

File T6col.cpli contains the row scores
 --- It has 37 rows and 1 columns

Col.	Mini	Maxi
1	-1.149e+02	2.992e+01



Lines

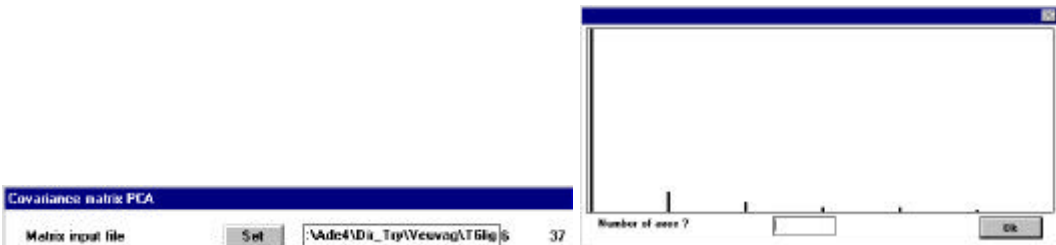
X file (default = 1, 2, 3, ..., n)

X file column number (default)

Y file (no default) 4\Dir_Try\Veuvag\T6col.cpli

3

Le tableau T6lig subit le même sort :



4

Centered Principal Component Analysis (Pearson 1901)
 Input file: T6lig
 ---- Row weights:
 File T6lig.cppl contains the row weights. It has 6 rows and 1 column. Each row has 1.6667e-01 weight (Sum = 1)

---- Column weights:
 File T6lig.cppc contains the column weights. It has 6 rows and 1 column. Each column has unit weight (Sum = 37)

---- Table:
 File T6lig.cpta contains the (column) centred table
 It has 6 rows and 37 columns

File :T6lig.cpta

Col.	Mini	Maxi
1	-4.400e-01	6.200e-01
2	-4.467e-01	6.533e-01
...		
36	-1.180e+01	5.888e+00
37	-4.705e+00	5.015e+00

---- Info: means and variances
 File T6lig.cpma contains the descriptive of the analysis
 It contains successively:

Number of rows: 6
 Number of columns: 37
 means and variances:
 Col.: 1 | Mean: 1.4400e+00 | Variance: 1.4453e-01
 Col.: 2 | Mean: 1.5967e+00 | Variance: 1.5039e-01
 ...
 Col.: 36 | Mean: 4.5712e+01 | Variance: 3.6676e+01
 Col.: 37 | Mean: 6.0575e+01 | Variance: 1.4882e+01

 DiagoRC: General program for two diagonal inner product analysis
 Input file: T6lig.cpta

--- Number of rows: 6, columns: 37

 Total inertia: 244.815

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.0617E+02	+0.8422	+0.8422	02	+2.1328E+01	+0.0871	+0.9293
03	+9.3449E+00	+0.0382	+0.9675	04	+4.0489E+00	+0.0165	+0.9840
05	+3.9194E+00	+0.0160	+1.0000	06	+0.0000E+00	+0.0000	+1.0000

File T6lig.cvpv contains the eigenvalues and relative inertia for each axis

--- It has 37 rows and 2 columns

File T6lig.cpcp contains the column scores

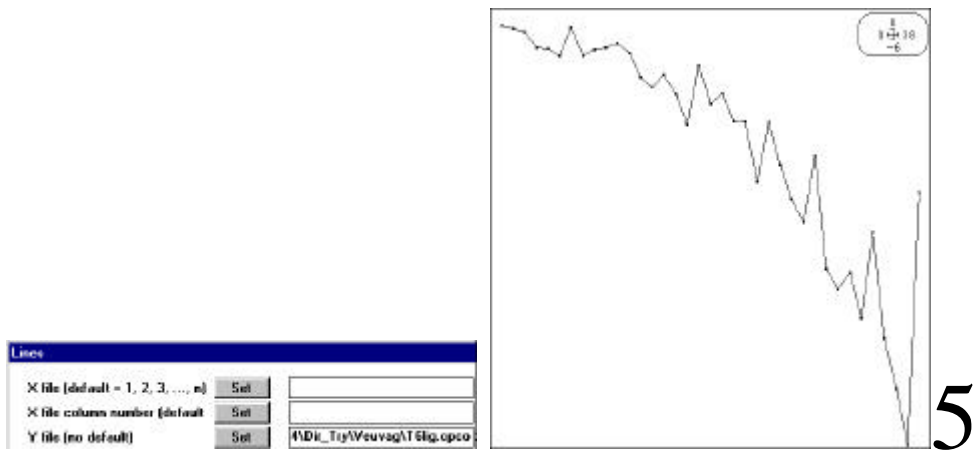
--- It has 37 rows and 1 columns

Col.	Mini	Maxi
1	-5.943e+00	-2.406e-01

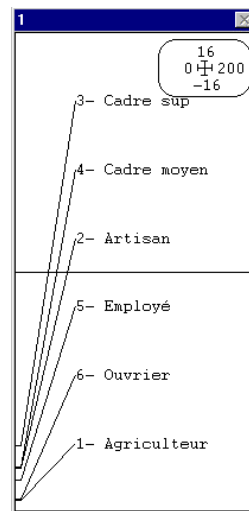
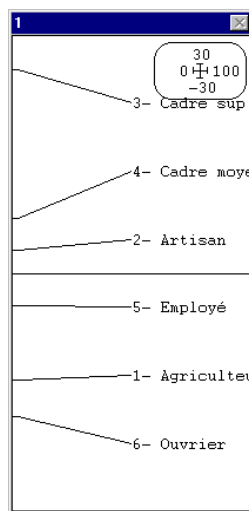
File T6lig.cpli contains the row scores

--- It has 6 rows and 1 columns

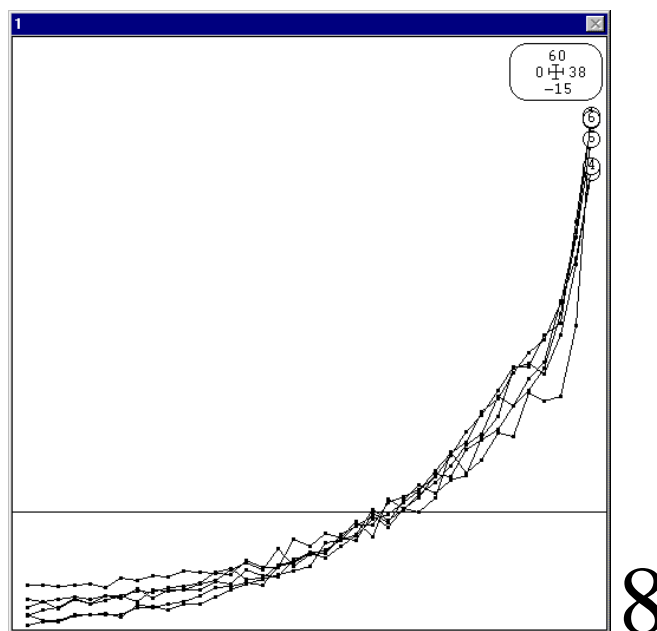
Col.	Mini	Maxi
1	-1.807e+01	2.568e+01



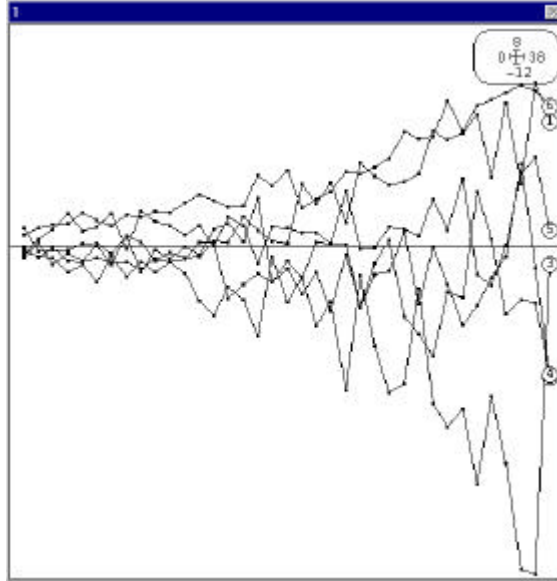
Dans les deux cas on a gardé un seul facteur. Dans les deux cas, on représente les catégories (Graph1D : Labels) :



Le cas n'est pas simple. La représentation des tableaux centrés s'impose. Pour l'un d'entre eux, c'est facile :



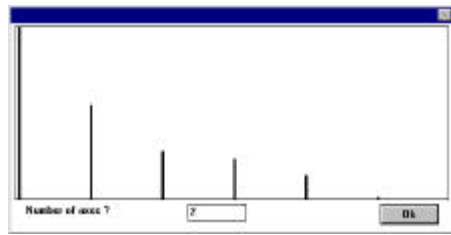
Pour l'autre, il faut une opération supplémentaire (Évidemment, il y a une question !) :



Tout s'éclaire. L'interprétation est facile. Par curiosité, on essaye l'analyse des correspondances :

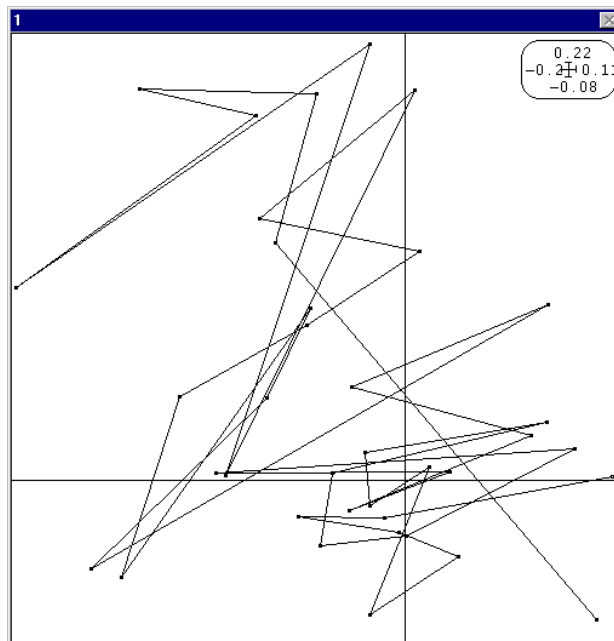
COrrespondence Analysis

Data file \Ade4\Dir_Try\Veuvag\T6col\37 6



Trajectories

XY coordinates file e4\Dir_Try\Veuvag\T6col.fcli 2



On aurait peut-être du regarder le listing :

```
fc/COA: Correspondence analysis
Input file: T6col
Number of rows: 37, columns: 6
File T6col.fcpl contains the margin distribution of rows
It has 37 rows and 1 column
File T6col.fcpc contains the margin distribution of columns
It has 6 rows and 1 column
File T6col.fcta contains the double centred table DI-1*P*DJ-1 -1I*1J'
It has 37 rows and 6 columns
File T6col.fcma contains:
    the number of rows: 37
    the number of columns: 6
    the total number: 3011.5
```

```
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: T6col.fcta
--- Number of rows: 37, columns: 6
-----
```

Total inertia: 0.00818776

```
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +3.7577E-03 +0.4589 +0.4589 | 02  +2.0464E-03 +0.2499 +0.7089 |
03  +1.0249E-03 +0.1252 +0.8341 | 04  +8.6158E-04 +0.1052 +0.9393 |
05  +4.9718E-04 +0.0607 +1.0000 | 06  +0.0000E+00 +0.0000 +1.0000 |
```

```
File T6col.fcvp contains the eigenvalues and relative inertia for each
axis
--- It has 6 rows and 2 columns
```

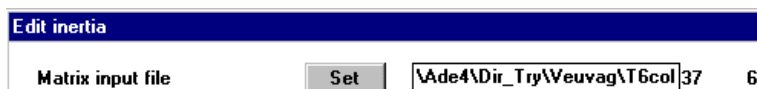
```
File T6col.fcco contains the column scores
--- It has 6 rows and 2 columns
```

```
File :T6col.fcco
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-5.028e-02| 1.487e-01|
|  2|-6.268e-02| 6.598e-02|
|----|-----|-----|
```

```
File T6col.fcli contains the row scores
--- It has 37 rows and 2 columns
```

```
File :T6col.fcli
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1|-1.975e-01| 1.052e-01|
|  2|-6.823e-02| 2.145e-01|
|----|-----|-----|
```

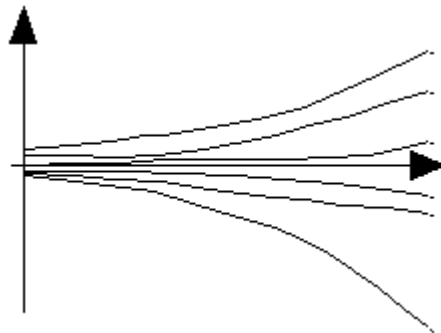
Un instant de réflexion est utile, peut-être. On ajoute simplement le listing de l'option HTA: Edit inertia, avec utilisation d'une pondération uniforme des lignes (1/37) et des colonnes (1/6) :



```
input file: D:\Ade4\Dir_Try\Veuvag\T6col
Option  1 = No action (non centred PCA)           Inertia = 2266.7
Option  2 = Centred table (overall centred PCA)   Inertia = 1162.61
Option  3 = Centred (zero mean) columns          Inertia = 1142.27
```

Option 4 = Centred (zero mean) rows	Inertia = 39.6998
Option 5 = Additive model	Inertia = 19.3656
Option 6 = Multiplicative model	Inertia = 10.9919

- 01** D'accord Pas d'accord ? La figure 3 exprime simplement une évidence.
- 02** D'accord Pas d'accord ? Les figures 3 et 6 sont associées de même que les figures 5 et 7.
- 03** D'accord Pas d'accord ? La figure 5 exprime simplement une autre évidence.
- 04** D'accord Pas d'accord ? Les deux ACP mettent en évidence un même fait observé.
- 05** D'accord Pas d'accord ? Il n'y a plus d'inégalité sociale devant la mort.
- 06** Vrai ou Faux ? La matrice des covariances de l'ACP de T6col est de dimension 6 et de rang 5.
- 07** Vrai ou Faux ? La matrice des covariances de l'ACP de T6lig est de dimension 37 et de rang 5.
- 08** D'accord Pas d'accord ? L'analyse des correspondances aurait été plus justifiée sur le tableau T6lig.
- 09** D'accord Pas d'accord ? On peut résumer les données par un schéma de principe du type ci-dessous.



- 10** D'accord Pas d'accord ? Les analyses linéaires ayant le même principe ont les mêmes fonctions.

4. Truites

Maîtrise BPE - UV MIAB II

Source : Lascaux, J.M. (1996) *Analyse de la variabilité morphologique de la truite commune (Salmo trutta L.) dans les cours d'eau du bassin pyrénéen méditerranéen*. Thèse de doctorat en sciences agronomiques, INP Toulouse. 1-160.

Dans 10 cours d'eau de quatre bassins hydrographiques, on a récolté 179 truites (*Salmo trutta*). On a mesuré sur chacune d'elle les variables suivantes :

PRAD	Nombre de points rouges avant l'aplomb de la Dorsale
PNAD	Nombre de points noirs avant l'aplomb de la Dorsale
PRAA	Nombre de points rouges après l'aplomb de l'Anale
PNAA	Nombre de points noirs après l'aplomb de l'Anale
PRDA	Nombre de points rouges entre aplomb Dorsale et aplomb Anale
PNDA	Nombre de points noirs entre aplomb Dorsale et aplomb Anale
PRLI	Nombre de points rouges sur la ligne latérale
PRINF	Nombre de points rouges au dessous de la ligne latérale
PNINF	Nombre de points noirs au dessous de la ligne latérale
PRSUP	Nombre de points rouges au dessus de la ligne latérale
PNSUP	Nombre de points noirs au dessus de la ligne latérale
PNO	Nombre de points noirs sur l'opercule
PRD	Nombre de points rouges sur la Dorsale
PND	Nombre de points noirs sur la Dorsale

L'analyse en composantes principales du tableau correspondant donne les valeurs propres :

```
Total inertia:      14
-----
```

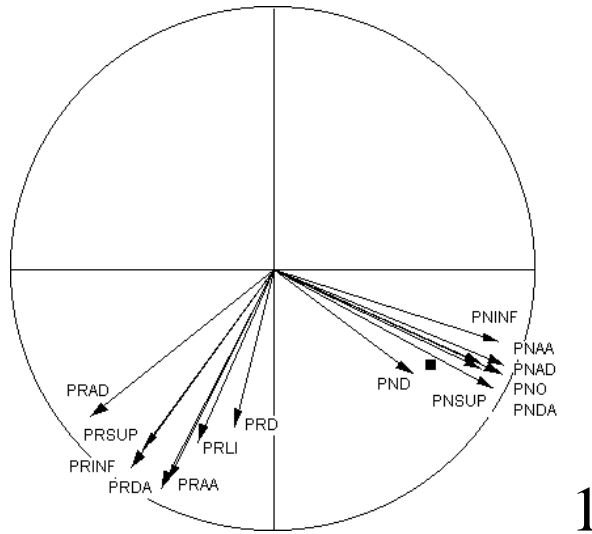
Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+6.0362E+00	+0.4312	+0.4312	02	+4.5386E+00	+0.3242	+0.7553
03	+7.1274E-01	+0.0509	+0.8063	04	+6.6982E-01	+0.0478	+0.8541
05	+6.2053E-01	+0.0443	+0.8984	06	+4.0095E-01	+0.0286	+0.9271
07	+2.9319E-01	+0.0209	+0.9480	08	+2.2580E-01	+0.0161	+0.9641
09	+1.8426E-01	+0.0132	+0.9773	10	+1.4323E-01	+0.0102	+0.9875
11	+9.9001E-02	+0.0071	+0.9946	12	+5.1393E-02	+0.0037	+0.9983
13	+1.6920E-02	+0.0012	+0.9995	14	+7.3662E-03	+0.0005	+1.0000

La matrice des corrélations est :

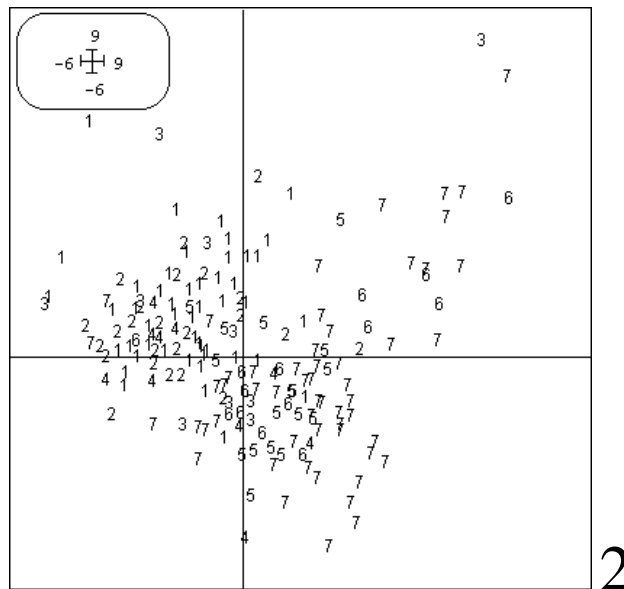
```
----- Correlation matrix -----
```

[1]	1000																		
[2]	-393	1000																	
[3]	630	-72	1000																
[4]	-329	773	-41	1000															
[5]	726	-56	810	-32	1000														
[6]	-359	914	-35	870	-34	1000													
[7]	572	-31	633	23	680	-8	1000												
[8]	805	-193	840	-147	897	-162	555	1000											
[9]	-451	921	-117	722	-105	855	-78	-220	1000										
[10]	722	-166	767	-153	800	-138	382	741	-243	1000									
[11]	-296	913	12	895	20	974	49	-113	794	-85	1000								
[12]	-320	842	-8	676	-54	777	33	-145	809	-157	767	1000							
[13]	417	79	470	33	466	93	352	485	27	399	107	161	1000						
[14]	-163	538	123	485	57	557	78	-27	432	53	577	627	133	1000					

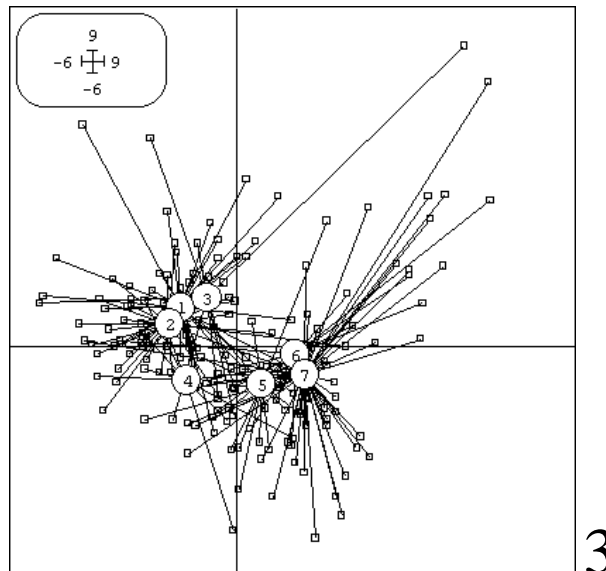
On peut classer ces truites en 7 groupes génétiques par le nombre d'allèles méditerranéens. La variable génétique prend les modalités 1 (0 allèle méditerranéen, homozygotes atlantiques, truites dites modernes), 2 à 6 (respectivement 1 à 5 allèles méditerranéens) et 7 (6 allèles méditerranéens, homozygotes méditerranéens, truites dites ancestrales).



Sur la carte des variables (plan 1-2) on a représenté cette variable supplémentaire par un point.



Sur la carte des individus dans le plan 1-2 on a représenté cette variable supplémentaire par des étiquettes.



Sur la carte des individus dans le plan 1-2 on a représenté cette variable supplémentaire par des centres de gravité des classes. Sur les mêmes truites on a mesuré 7 variables méristiques :

RD	Nombre de rayons à la dorsale
RA	Nombre de rayons à l'anale
RPELG	Nombre de rayons à la pelvienne gauche
RPECG	Nombre de rayons à la pectorale gauche
BR	Nombre de branchiospines (1°arc gauche)
VERT	Nombre de vertèbres
CAEC	Nombre de caeca pyloriques

L'analyse en composantes principales du tableau correspondant donne le listing :

```
Classical Principal Component Analysis (Hotelling 1933)
Input file: merist
---- Row weight:
File merist.cnpl contains the row weight
It has 179 rows and 1 column
Each row has 5.5866e-03 weight (Sum = 1)
---- Column weights:
File merist.cnpc contains the column weights
It has 7 rows and 1 column
Each column has unit weight (Sum = 7)
---- Table:
File merist.cnta contains the centred and normed table
Zero mean and unit variance for each column
It has 179 rows and 7 columns
```

```
File :merist.cnta
|Col.|  Mini  |  Maxi  |
|----|-----|-----|
|  1  | -3.243e+00 | 1.705e+00 |
|  2  | -4.111e+00 | 1.367e+00 |
|  3  | -5.087e+00 | 5.031e+00 |
|  4  | -3.698e+00 | 1.474e+00 |
|  5  | -2.579e+00 | 2.758e+00 |
|  6  | -1.884e+00 | 1.863e+00 |
|  7  | -2.065e+00 | 2.717e+00 |
|----|-----|-----|
```



```

---- Info: means and variances
File merist.cnma contains the descriptive of the analysis
It contains successively:
  Number of rows: 179
  Number of columns: 7
  means and variances:
  Col.: 1 | Mean: 1.0966e+01 | Variance: 3.6759e-01
  Col.: 2 | Mean: 9.2514e+00 | Variance: 2.9993e-01
  Col.: 3 | Mean: 9.0056e+00 | Variance: 3.9075e-02
  Col.: 4 | Mean: 1.3145e+01 | Variance: 3.3644e-01
  Col.: 5 | Mean: 1.7899e+01 | Variance: 1.2636e+00
  Col.: 6 | Mean: 5.8508e+01 | Variance: 6.4099e-01
  Col.: 7 | Mean: 4.6726e+01 | Variance: 1.0077e+02

```

```

-----
File merist.cn+r contains the Correlation matrix
from statistical triplet merist.cnta
It has 7 rows and 7 columns
----- Correlation matrix -----

```

```

[ 1] 1000
[ 2] 227 1000
[ 3] 48 -116 1000
[ 4] 62 166 90 1000
[ 5] -161 50 3 237 1000
[ 6] 127 65 -53 82 -24 1000
[ 7] -55 -149 153 41 147 -83 1000

```

```

-----
DiagoRC: General program for two diagonal inner product analysis
Input file: merist.cnta
--- Number of rows: 179, columns: 7
-----
Total inertia: 7
-----

```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.4518E+00	+0.2074	+0.2074	02	+1.3403E+00	+0.1915	+0.3989
03	+1.1164E+00	+0.1595	+0.5584	04	+9.4451E-01	+0.1349	+0.6933
05	+8.3484E-01	+0.1193	+0.8126	06	+6.7703E-01	+0.0967	+0.9093
07	+6.3503E-01	+0.0907	+1.0000				

```

File merist.cnvp contains the eigenvalues and relative inertia for each
axis
--- It has 7 rows and 2 columns

```

```

File merist.cnco contains the column scores
--- It has 7 rows and 2 columns
File :merist.cnco
|Col.| Mini | Maxi |
|----|-----|-----|
| 1|-5.437e-01| 6.481e-01|
| 2| 6.384e-02| 7.742e-01|
|----|-----|-----|

```

```

File merist.cnli contains the row scores
--- It has 179 rows and 2 columns
File :merist.cnli
|Col.| Mini | Maxi |
|----|-----|-----|
| 1|-4.886e+00| 2.788e+00|
| 2|-3.334e+00| 4.041e+00|
|----|-----|-----|

```

L'analyse de variance de chacune des variables pour la répartition en groupes génétiques donne le listing :

variable 1 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	3.634	6	0.6057	1.676	0.1286
Within	62.16	172	0.3614		
Total	65.8	178			

variable 2 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	3.904	6	0.6507	2.248	0.04065
Within	49.78	172	0.2894		
Total	53.69	178			

variable 3 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	0.295	6	0.04917	1.262	0.2766
Within	6.699	172	0.03895		
Total	6.994	178			

variable 4 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	0.3809	6	0.06348	0.1824	0.9798
Within	59.84	172	0.3479		
Total	60.22	178			

variable 5 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	5.17	6	0.8616	0.6705	0.6756
Within	221	172	1.285		
Total	226.2	178			

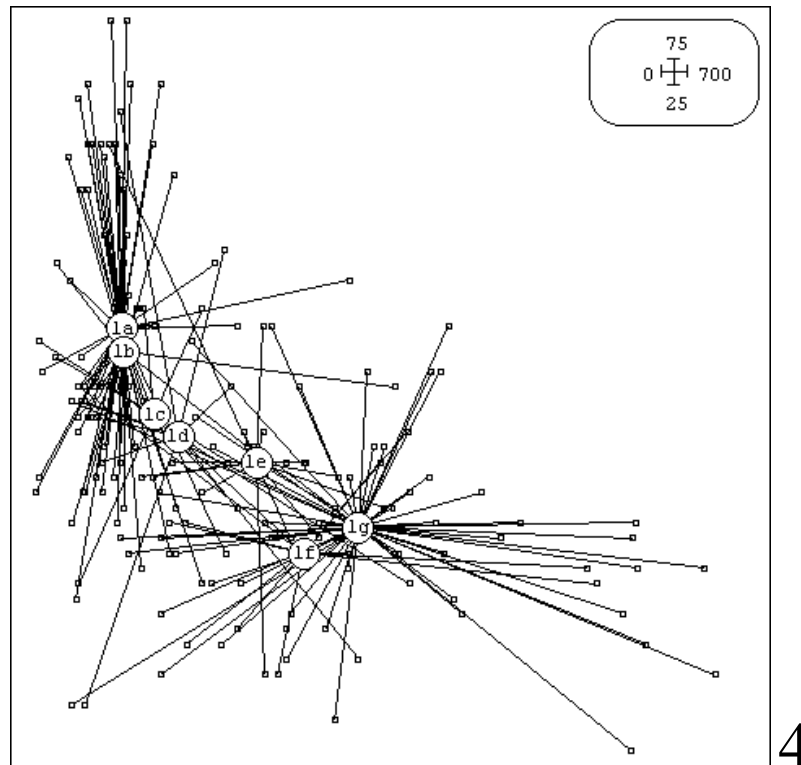
variable 6 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	2.537	6	0.4229	0.6483	0.6935
Within	112.2	172	0.6523		
Total	114.7	178			

variable 7 from merist versus variable 1 from Groupe

Source	SS	d.f.	MS	F	Proba
Between	6095	6	1016	14.63	0
Within	1.194E+04	172	69.43		
Total	1.804E+04	178			

On ajoute le graphe 4 (en abscisse, le nombre total de points noirs ; en ordonnée, le nombre de caeca pyloriques). Les lettres désignent le groupe génétique.



01 D'accord Pas d'accord ? La coloration de la robe des truites est vraisemblablement sous la dépendance de deux mécanismes indépendants.

02 D'accord Pas d'accord ? Les variables méristiques sont sans déterminisme génétique.

03 D'accord Pas d'accord ? D'accord Pas d'accord ? Il est logique de ne dépouiller que les deux premiers axes dans la première analyse.

04 D'accord Pas d'accord ? La seconde ACP ne nous apprend rien.

05 D'accord Pas d'accord ? Les fortes corrélations entre variables de coloration sont en partie artéfactuelles.

06 D'accord Pas d'accord ? Les variables mesurées sont suffisantes pour substituer l'approche morphologique à l'analyse génétique.

5. Kyste de l'ovaire

DEA AMSB / Décembre 1997

Source :Lasne (Y.), Akiki (S.), Dargent (D.) 1988 Dosages des marqueurs biochimiques dans les kystes de l'ovaire. Communication au 7ème Colloque International de Biologie Prospective. Pont-à-Mousson.

Dargent (D.), Lasne (Y.) & Akiki (S.) (1990). Diagnostic de la nature des kystes de l'ovaire par le dosage dans le liquide kystique et dans le sang des stéroïdes et des marqueurs tumoraux". Dédutions pratiques. *Contraception-fertilité-sexualité*, 18, 11, 1011-1017.

La question est posée par D. Dargent : y a-t'il un ou des marqueurs biochimiques capables de rendre compte de la fonctionnalité, de l'organicité et au sein de celle-ci de la bénignité ou de la malignité d'un kyste ovarien ? 126 patientes présentant un kyste de l'ovaire constituent un échantillon de référence pour comparer diagnostic histologique et diagnostic biochimique, l'intérêt manifeste du second étant d'éviter une intervention chirurgicale qui ne s'imposerait pas. Le diagnostic histologique met en évidence 76 cas de kystes organiques bénins (KOB), 25 cas de cancers (kystes organiques malins KOM) et 25 cas de kystes fonctionnels (KF). Avant l'opération une ponction de liquide du kyste est effectuée simultanément avec une prise de sang. Dans le sérum et le liquide kystique sont dosés 5 molécules, l'antigène carcino-embryonnaire (ACE en ng/ml), le CA 19-9 et le CA 125 (ui/ml), l'oestradiol (E) et la progestérone (P) (ng/ml). On obtient ainsi 10 variables (1 à 5 dosages sérum et 6 à 10 dosages liquide) mesurées sur 126 patientes réparties en 3 groupes.

I- Analyse univariée

Les données brutes sont transformées par le changement de variable $y = \text{Log}(x+1)$. Les données transformées donnent les moyennes et les variances :

VAR: 1	MOY:+.1198E+01	VAR:+.5607E-01	(ACE, sérum)
VAR: 2	MOY:+.2820E+01	VAR:+.7579E+00	(CA 19-9, sérum)
VAR: 3	MOY:+.3333E+01	VAR:+.2622E+01	(CA 125, sérum)
VAR: 4	MOY:+.1476E+00	VAR:+.1262E+00	(œstradiol, sérum)
VAR: 5	MOY:+.1091E+01	VAR:+.7694E+00	(progestérone, sérum)
VAR: 6	MOY:+.2130E+01	VAR:+.3518E+01	(ACE, kyste)
VAR: 7	MOY:+.5286E+01	VAR:+.1409E+02	(CA 19-9, kyste)
VAR: 8	MOY:+.7321E+01	VAR:+.1357E+02	(CA 125, kyste)
VAR: 9	MOY:+.1025E+01	VAR:+.3528E+01	(œstradiol, kyste)
VAR: 10	MOY:+.2112E+01	VAR:+.6245E+01	(progestérone, kyste)

Les données transformées sont consignées dans un fichier DOL (Dosage-Log) de 126 lignes et 10 colonnes. Une analyse de variance à un facteur contrôlé (le diagnostic histologique) est effectuée sur chacune des variables. On obtient :

```
variable 1 from DOL versus variable 1 from Diagnostic
-----
```

Source		SS d.f.		MS		F		Proba
--------	--	---------	--	----	--	---	--	-------

Between	0.1906	2	0.0953	1.705	0.1839
Within	6.875	123	0.05589		
Total	7.065	125			

 variable 2 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	6.782	2	3.391	4.702	0.01077
Within	88.71	123	0.7212		
Total	95.49	125			

 variable 3 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	247.6	2	123.8	183.8	0
Within	82.84	123	0.6735		
Total	330.4	125			

 variable 4 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	0.0988	2	0.0494	0.3844	0.6871
Within	15.81	123	0.1285		
Total	15.91	125			

 variable 5 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	6.693	2	3.346	4.561	0.01224
Within	90.25	123	0.7338		
Total	96.95	125			

 variable 6 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	27.27	2	13.63	4.031	0.01983
Within	416	123	3.382		
Total	443.3	125			

 variable 7 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	306.2	2	153.1	12.82	1.586E-05
Within	1469	123	11.94		
Total	1775	125			

 variable 8 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	919	2	459.5	71.42	0
Within	791.3	123	6.434		
Total	1710	125			

 variable 9 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	292.2	2	146.1	118	0
Within	152.3	123	1.238		
Total	444.5	125			

variable 10 from DOL versus variable 1 from Diagnostic

Source	SS	d.f.	MS	F	Proba
Between	487.4	2	243.7	100.1	0
Within	299.5	123	2.435		
Total	786.9	125			

Les variables sont normalisées. Les moyennes par groupes sont :

	1	2	3
1	-0.02	0.29	-0.23
2	-0.11	0.53	-0.20
3	-0.42	1.74	-0.48
4	0.05	0.02	-0.16
5	0.12	-0.53	0.15
6	0.13	0.12	-0.50
7	0.25	0.06	-0.82
8	0.24	0.70	-1.43
9	-0.43	-0.31	1.63
10	-0.35	-0.52	1.58

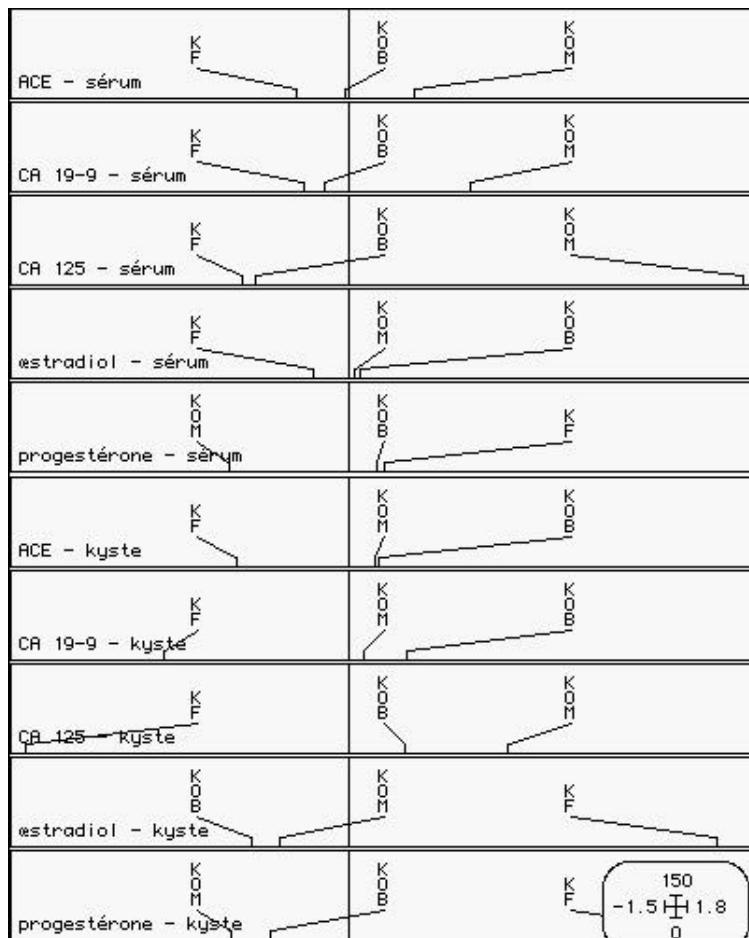


Figure 1

II Analyse en composantes principales

L'analyse en composantes principales normée du tableau DOL donne :

Classical Principal Component Analysis (Hotelling 1933)

Input file: DOL

---- Row weight:

File DOL.cnpl contains the row weight

It has 126 rows and 1 column

Each row has 7.9365e-03 weight (Sum = 1)

---- Column weights:

File DOL.cnpc contains the column weights

It has 10 rows and 1 column

Each column has unit weight (Sum = 10)

---- Table:

File DOL.cnta contains the centred and normed table

Zero mean and unit variance for each column

It has 126 rows and 10 columns

File :DOL.cnta

Col.	Mini	Maxi
1	-4.203e-01	5.759e+00

...

10	-8.453e-01	3.042e+00
----	------------	-----------

File DOL.cn+r contains the Correlation matrix

from statistical triplet DOL.cnta

It has 10 rows and 10 columns

----- Correlation matrix -----

[1]	1000
[2]	88 1000
[3]	137 285 1000
[4]	2 270 59 1000
[5]	22 -44 -259 562 1000
[6]	38 67 87 334 235 1000
[7]	21 300 126 380 220 629 1000
[8]	91 112 359 28 -116 173 286 1000
[9]	34 -39 -163 35 141 -146 -260 -554 1000
[10]	-89 -103 -260 135 342 -57 -188 -539 857 1000

Total inertia: 10

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.7877E+00	+0.2788	+0.2788	02	+2.2295E+00	+0.2229	+0.5017
03	+1.2579E+00	+0.1258	+0.6275	04	+9.6660E-01	+0.0967	+0.7242
05	+8.3566E-01	+0.0836	+0.8077	06	+7.1325E-01	+0.0713	+0.8790
07	+4.7350E-01	+0.0474	+0.9264	08	+3.2780E-01	+0.0328	+0.9592
09	+3.0181E-01	+0.0302	+0.9894	10	+1.0640E-01	+0.0106	+1.0000

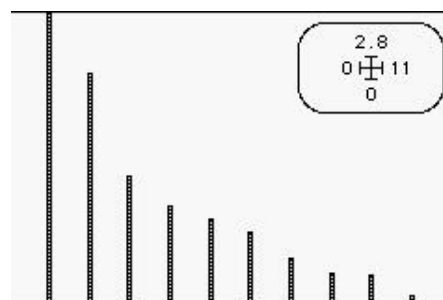


Figure 2

File DOL.cnco contains the column scores
 --- It has 10 rows and 2 columns

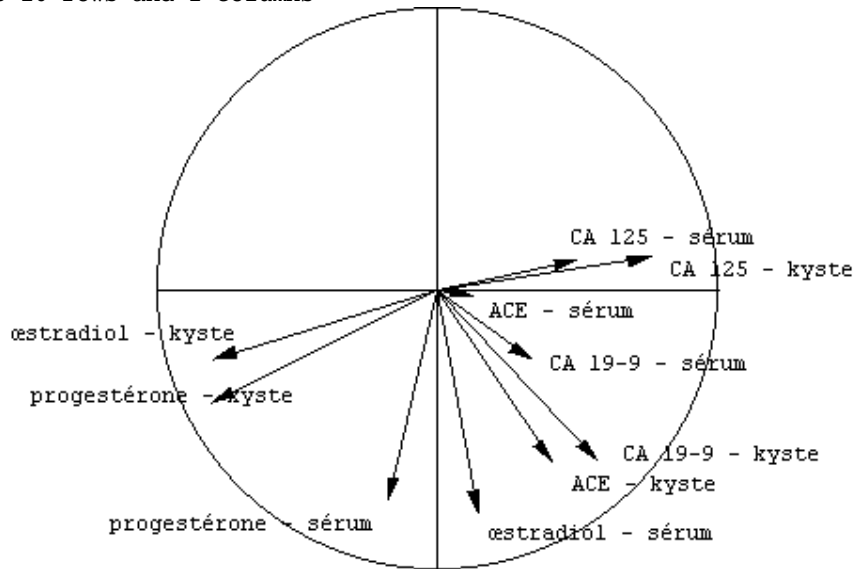


Figure 3

File DOL.cnli contains the row scores
 --- It has 126 rows and 2 columns

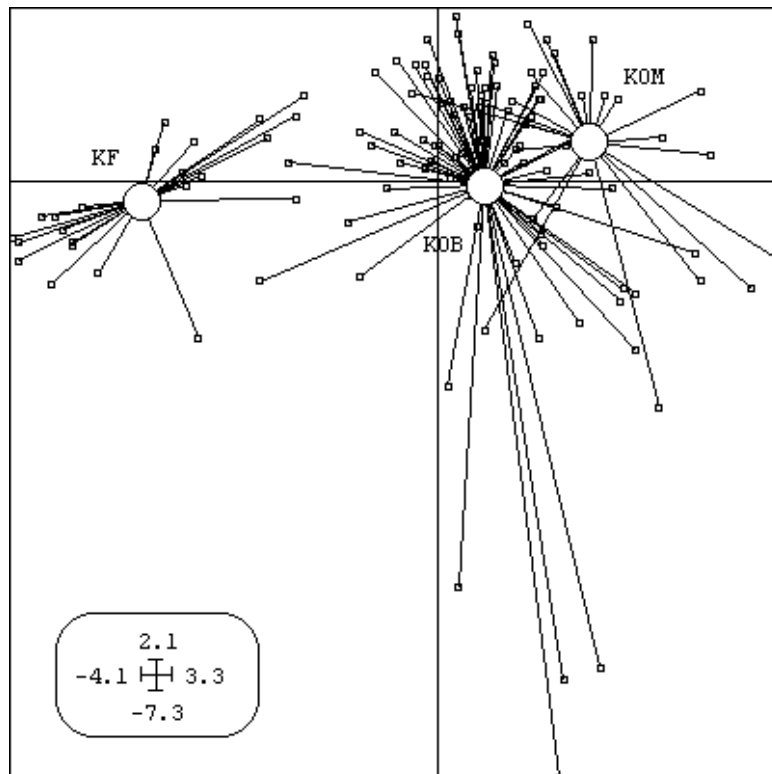


Figure 4

III ACP inter-classe

2	-1.937e+00	3.085e+00
-----	-----	-----

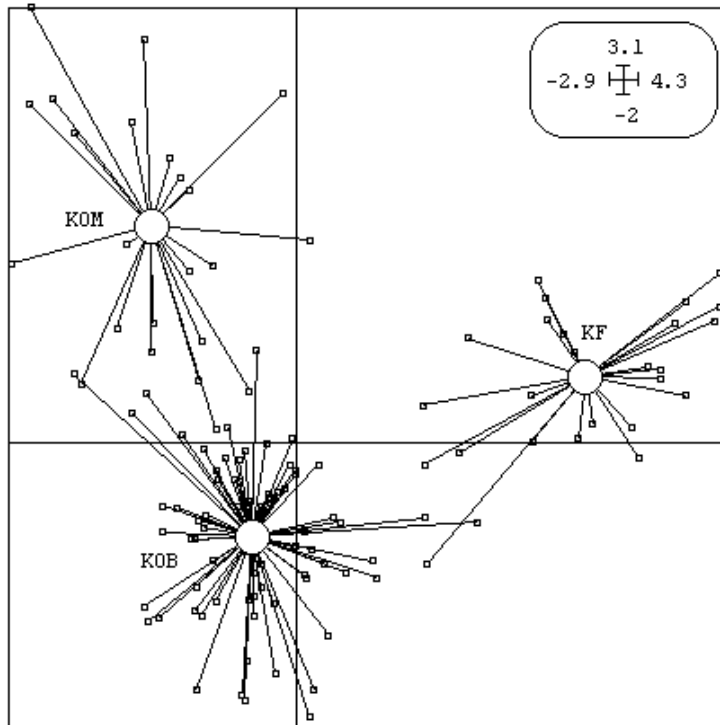


Figure 7

File DD.beco contains the standard column scores (lambda norm)
It has 10 rows and 2 columns

File :DD.beco

Col.	Mini	Maxi
1	-7.328e-01	7.743e-01
2	-2.222e-01	7.403e-01

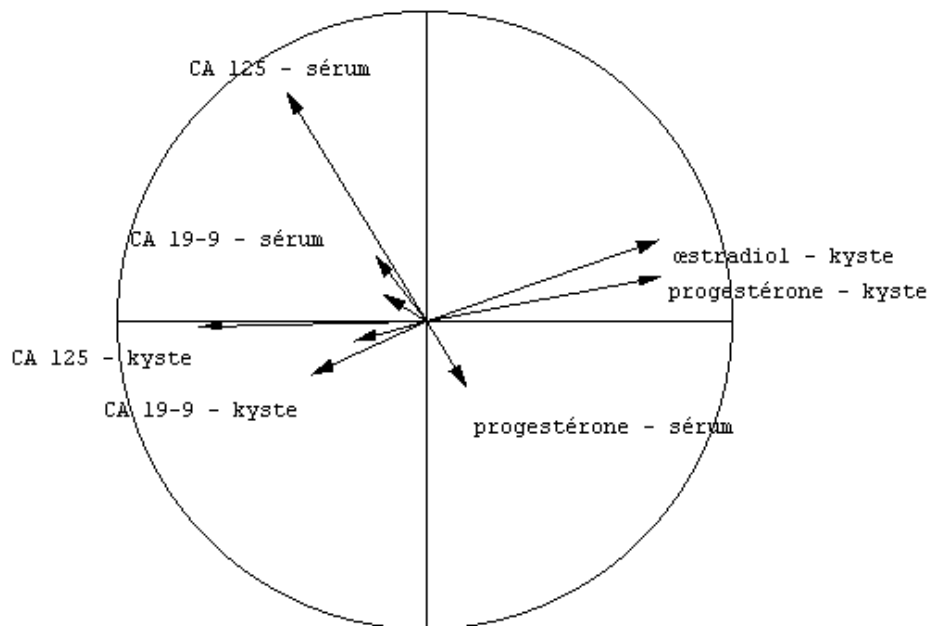


Figure 8

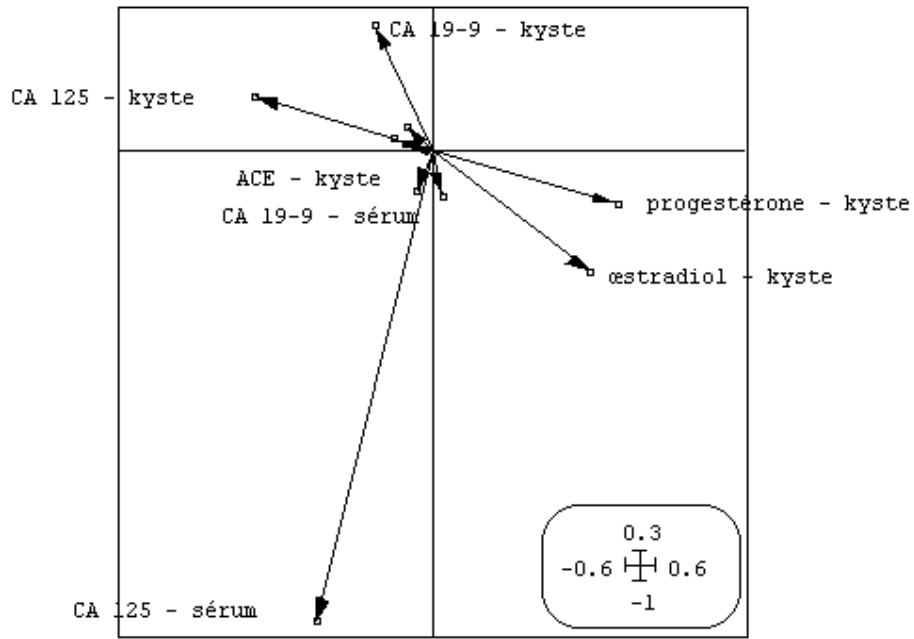


Figure 10

File DD.diap contains the principal axes

It has 10 rows and 2 columns

File :DD.diap

Col.	Mini	Maxi
1	-8.111e-01	8.581e-01
2	-8.592e-01	2.580e-01

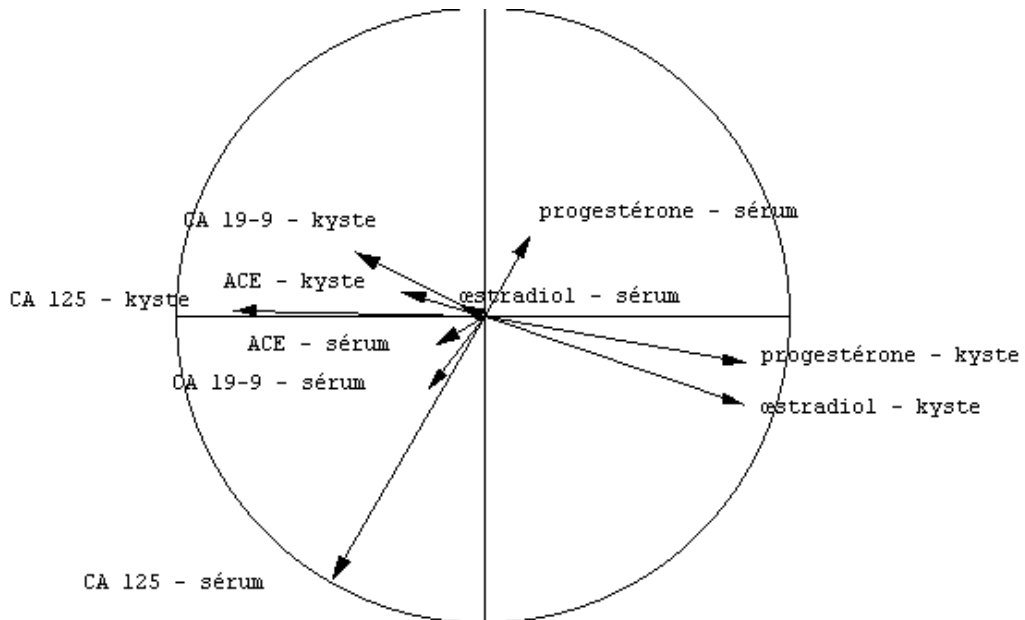


Figure 11

File DD.dili contains the canonical scores of row (unit norm)

It has 126 rows and 2 columns

File :DD.dili

Col.	Mini	Maxi
1	-1.823e+00	2.677e+00

2	-3.439e+00	1.443e+00
-----	-----	-----

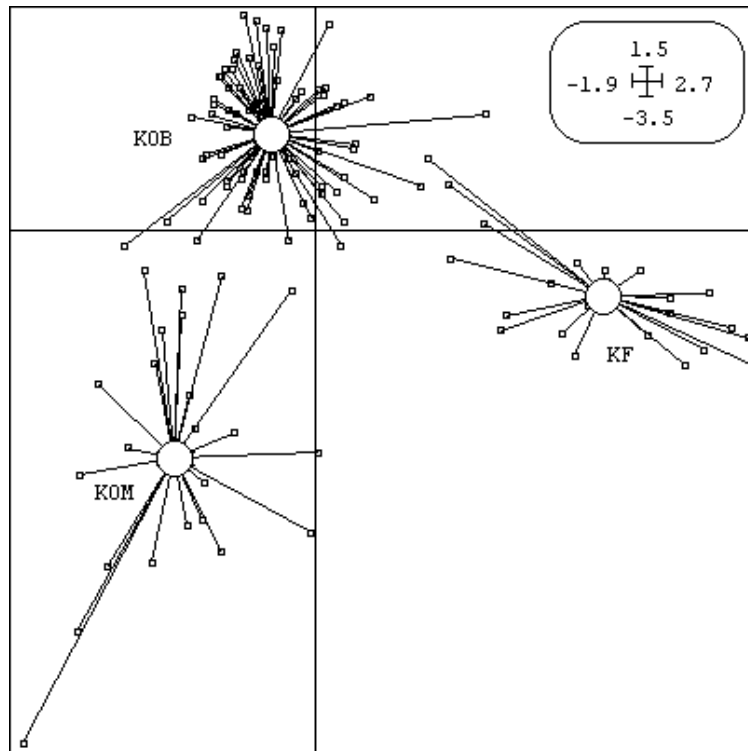


Figure 12

File DD.dicp contains the correlations between PCA scores and DA scores. It has 10 rows and 2 columns

File :DD.dicp

Col.	Mini	Maxi
1	-9.678e-01	1.042e-01
2	-7.542e-01	1.763e-01

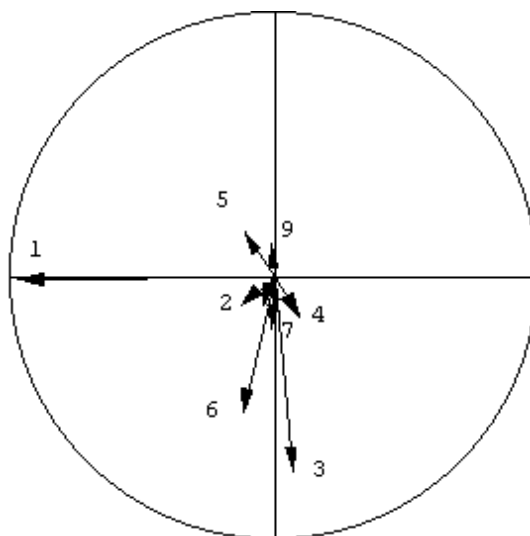


Figure 13



Figure 14

Entrée des résultats des dosages
(Attention aux unités)

	SÉRUM	KYSTE
R.E.C. (ng/ml)	2.80	2.00
C.A. 19-9 (ui/ml)	12.0	104833.00
C.A. 125 (ui/ml)	46.92	24329.00
Testostérone (ng/ml)	0.01	0.03
Progesterone (ng/ml)	4.19	17.79

Valider

Figure 15

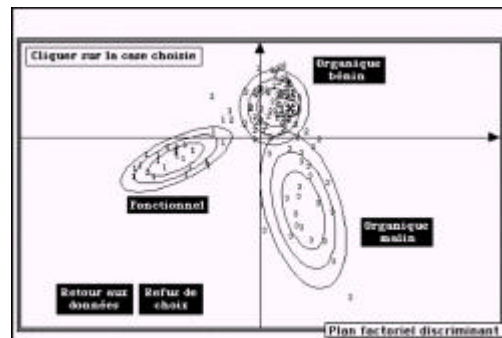


Figure 16

- 01 D'accord Pas d'accord ? L'analyse statistique univariée suffit à atteindre l'objectif principal de l'expérience.
- 02 D'accord Pas d'accord ? La statistique univariée ne suffit pas et l'utilisation de l'analyse en composantes principales est ici très utile.
- 03 D'accord Pas d'accord ? Dans plus de 90% le diagnostic biochimique est possible et valide.
- 04 D'accord Pas d'accord ? On peut réduire le nombre de molécules étudiées.
- 05 D'accord Pas d'accord ? Il est possible d'éviter aux patientes le prélèvement de liquide kystique.
- 06 D'accord Pas d'accord ? Les deux principales valeurs de la matrice de corrélation ont la même origine.
- 07 D'accord Pas d'accord ? Le pourcentage de variance interclasse de la première colonne du fichier DD.bels ne peut dépasser 82 % et sa variance elle-même ne peut dépasser 2.79.
- 08 D'accord Pas d'accord ? L'axe 2 de l'analyse discriminante est voisin de l'axe 4 de l'ACP initiale .
- 09 D'accord Pas d'accord ? Les données ayant permis la construction du logiciel GRAPH-KYSTE sont celles qui sont décrites dans les calculs.

10 D'accord Pas d'accord ? Les triplets statistiques utilisés en ACP normée et en analyse discriminante ont un élément commun

11 D'accord Pas d'accord ? Les triplets statistiques utilisés en ACP interclasses et en analyse discriminante ne diffèrent que par un seul élément.