



HAL
open science

Analyses Multifactorielles : Analyses en Composantes principales

Sandrine Mignon-Grasteau

► **To cite this version:**

Sandrine Mignon-Grasteau. Analyses Multifactorielles : Analyses en Composantes principales. Master. MASTER Sciences, Technologies, Santé MENTION Biologie Intégrative et Agrosociétés SPÉCIALITÉ Qualité et environnement en productions animales FINALITÉ Professionnelle (Statistiques), 2014. hal-02801559

HAL Id: hal-02801559

<https://hal.inrae.fr/hal-02801559>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSES DESCRIPTIVES

Sandrine Mignon-Grasteau

2014

TYPES D'ANALYSES

- **ACP** (Analyse en Composante Principale)
- **AFC** (Analyse Factorielle des Correspondances)
- **Analyses Discriminantes**
- **Méthodes de classification**
 - ↪ Décrire un jeu de données

ACP (1)

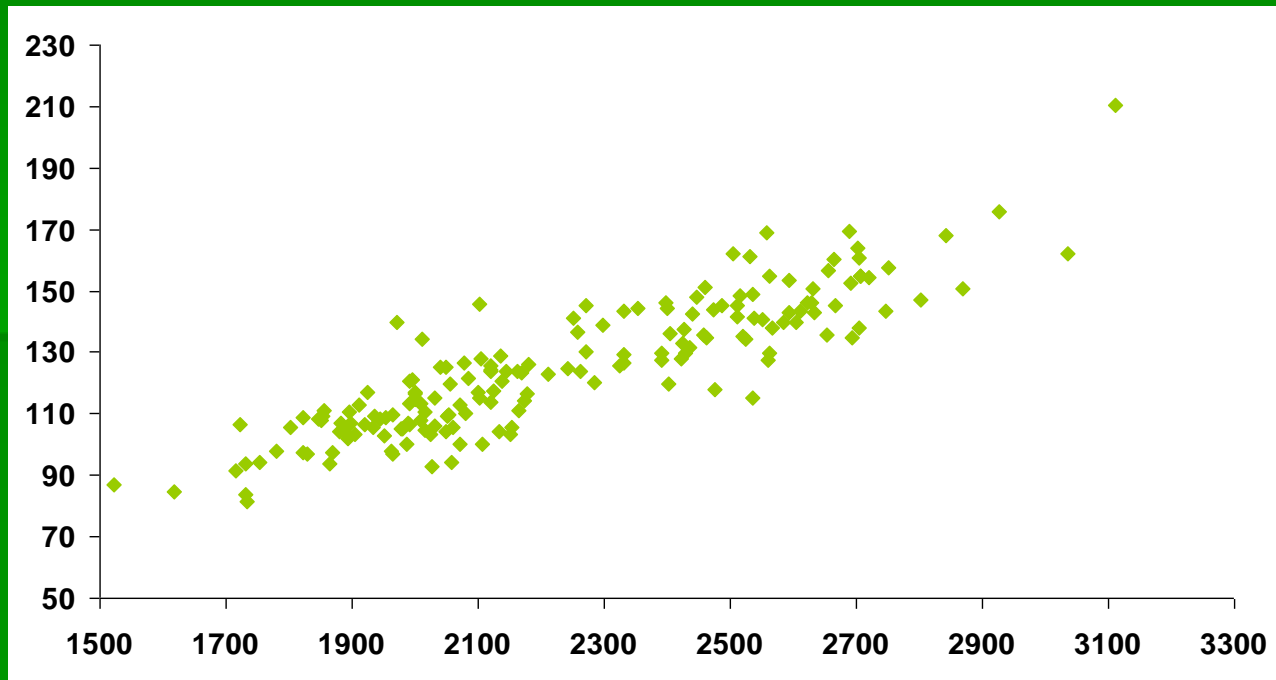
- Données quantitatives (poids, taille ...)
- 2 variables ou +, mesurées sur un nombre « confortable » d'individus*

* individu = ligne d'un tableau

ACP (2)

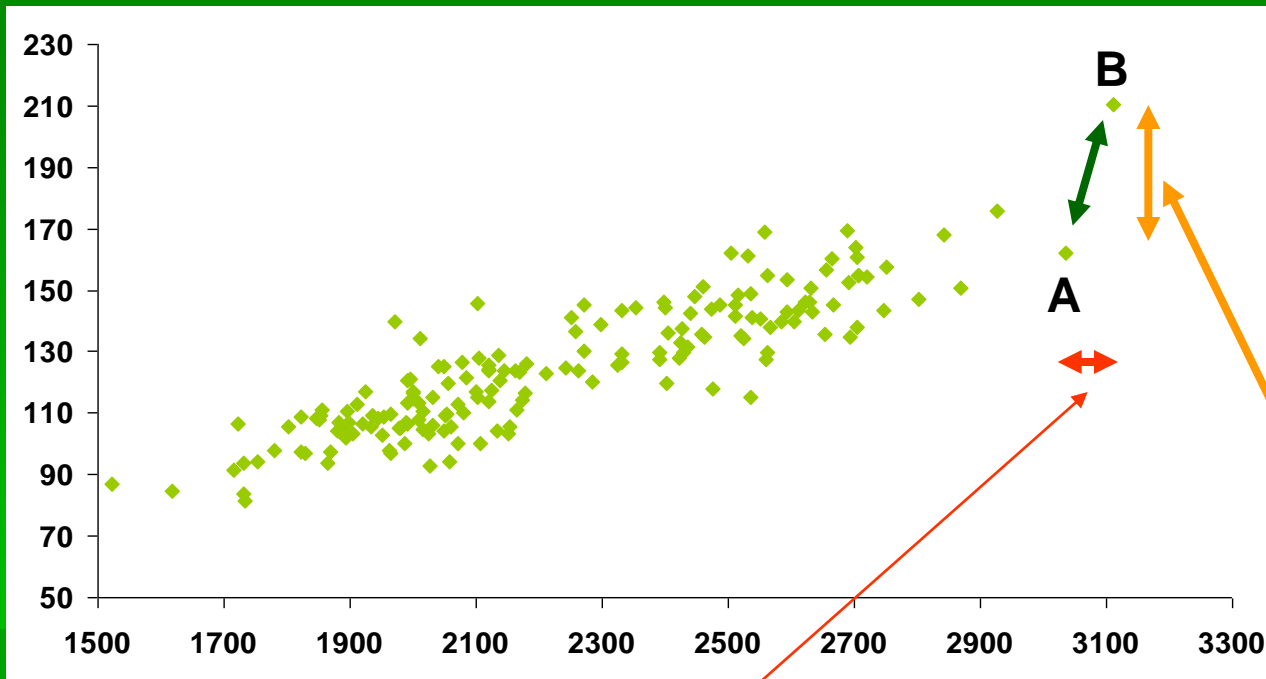
- Si seulement deux variables => représentation en « xy »

Poids du
filet (g)



Poids du poulet (g)

ACP (2) : distance entre 2 points

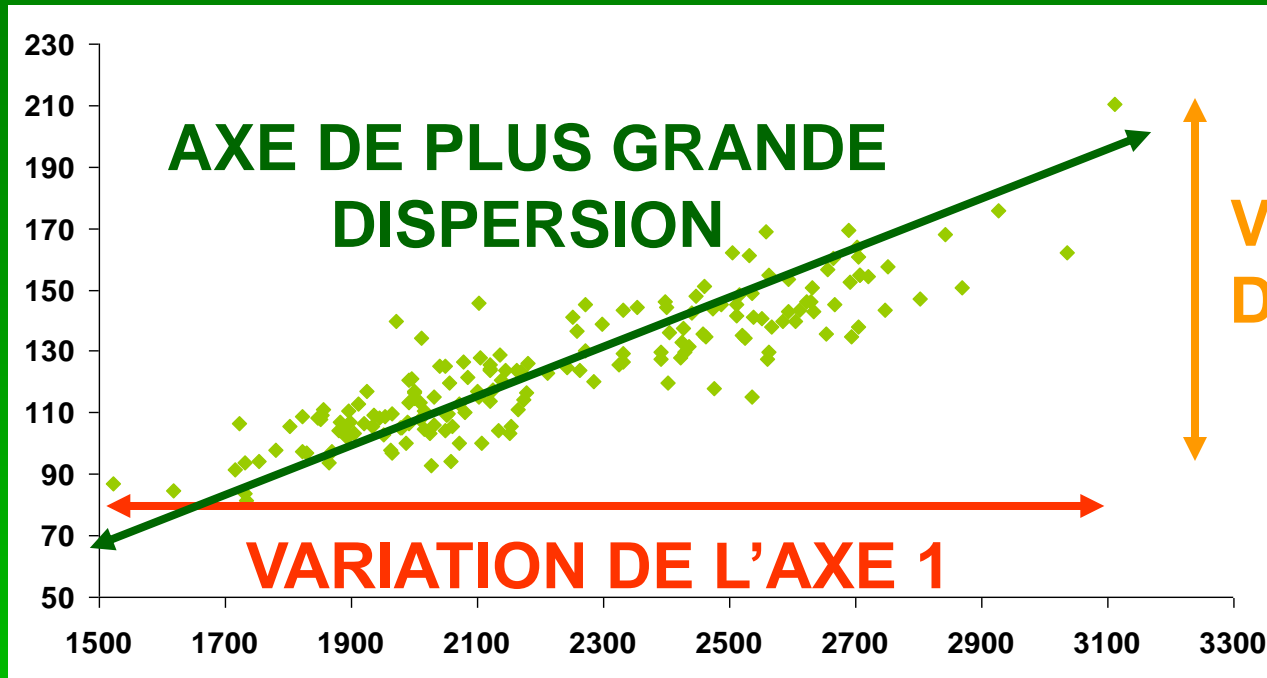


$$\text{Dist}_{A-B} = \sqrt{(\text{Poids}_A - \text{Poids}_B)^2 + (\text{Filet}_A - \text{Filet}_B)^2}$$

ACP (2) : INERTIE

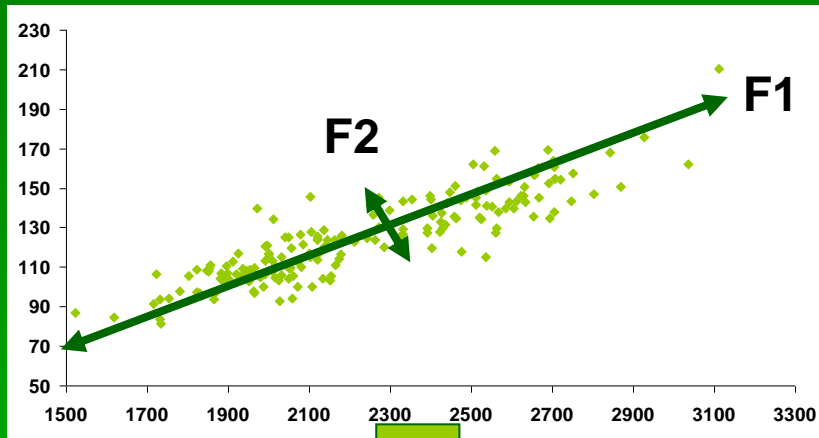
- Inertie =
 - « Variance totale du nuage de points »
 - Somme des variances de chaque variable ($\sigma^2(\text{Poids}) + \sigma^2(\text{Filet})$)
 - Somme des distances entre le centre de gravité du nuage et chacun des points
- **But de l'ACP** : décrire le nuage de points et trouver les sources de variation
- **Moyen** : Rechercher un axe décrivant le mieux possible la dispersion du nuage

ACP : Axe Factoriel



- Longueur de l'axe de plus grande dispersion = valeur propre
- Axe ayant la plus grande valeur propre =
 - Premier axe factoriel
 - Facteur 1 (F1)

ACP : Axe Factoriel

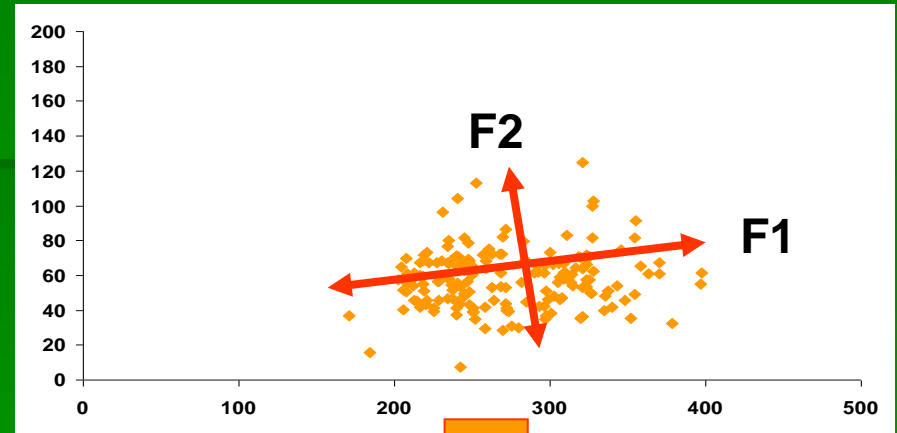


Le Facteur 1 rend compte d'un fort % de l'inertie

⇒ le 2^{ème} facteur n'explique pas grand chose

F2 :

- orthogonal à F1
- a la 2^{ème} plus grande valeur propre

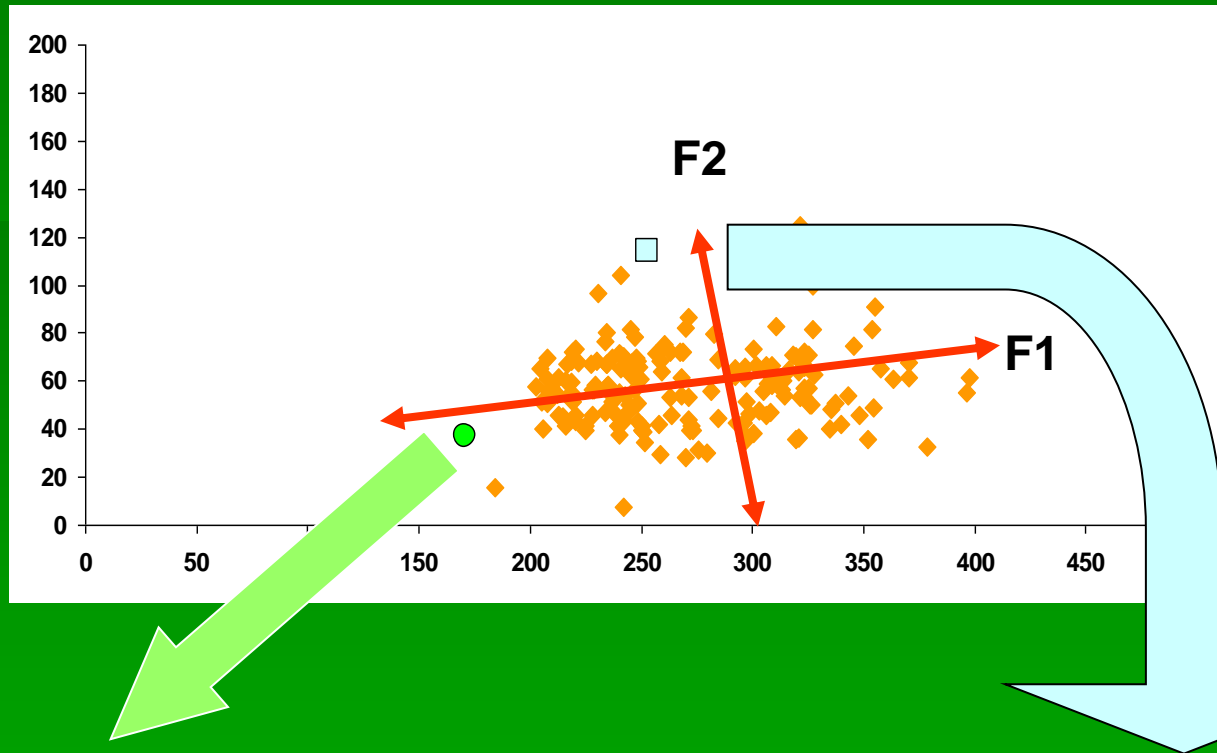


Le Facteur 1 rend compte d'un % modéré de l'inertie

⇒ le 2^{ème} facteur explique quelque chose

- passe par le centre de gravité
- explique un % d'inertie inférieur à F1

ACP : Contribution d'un point à un axe

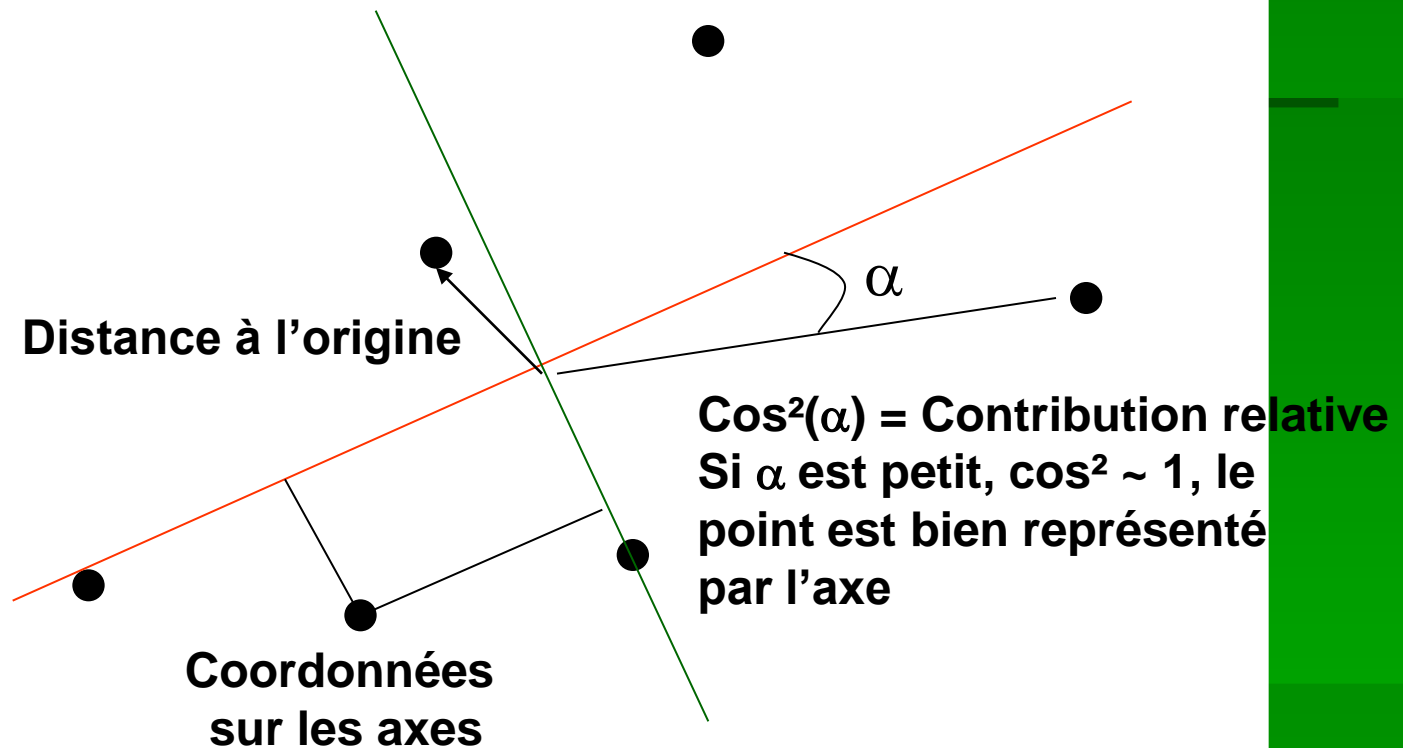


Ce point explique une proportion importante de la dispersion de F1, mais pas de F2

Ce point explique une proportion importante de la dispersion de F2, mais pas de F1

- ❖ Ces points ont des « contributions » différentes aux axes F1 et F2
- ❖ Pour chaque facteur, la somme des contributions de chaque point vaut 100%

ACP : paramètres de description d'un point

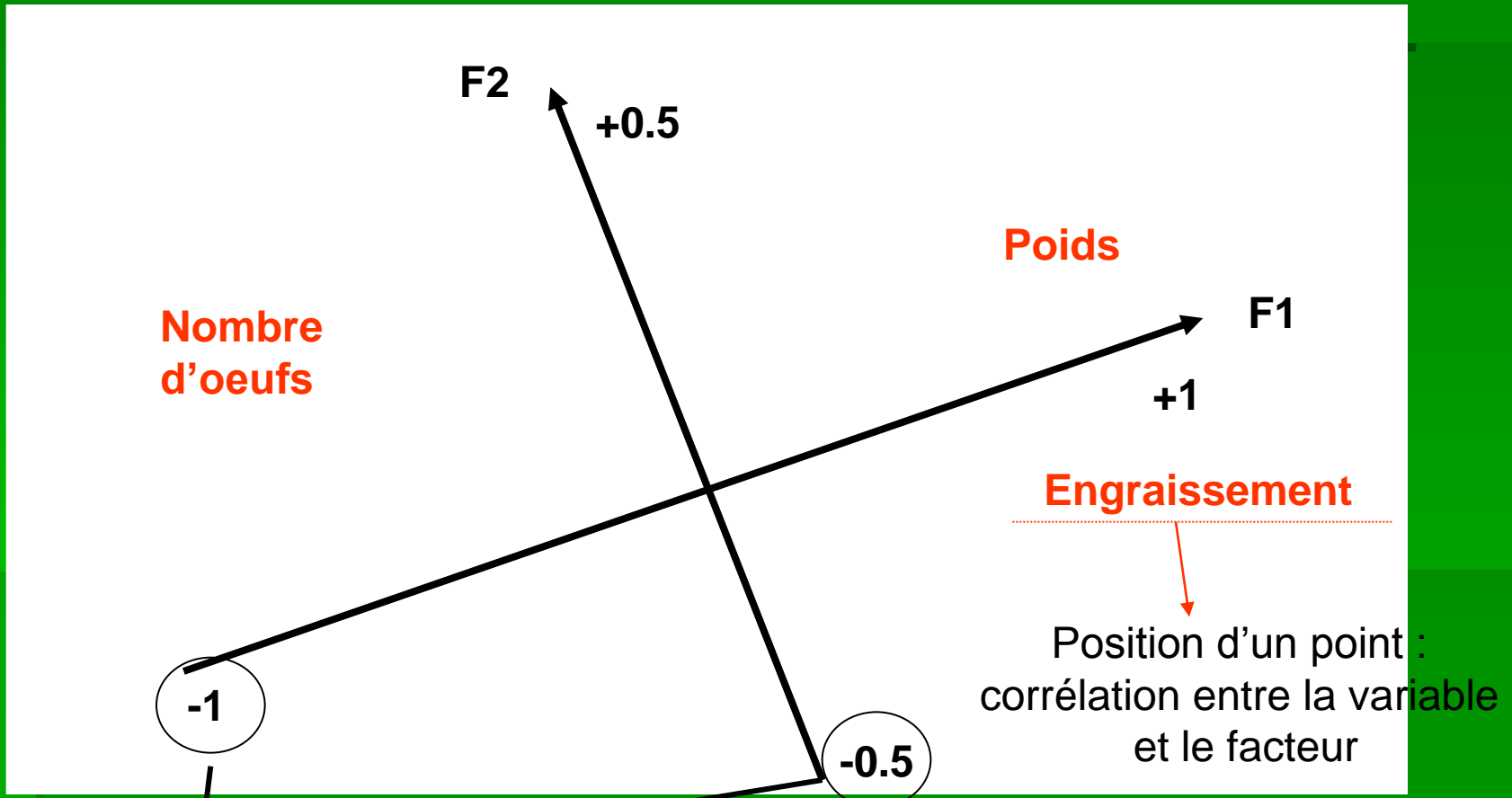


Contribution du point à l'axe : dans quelle mesure le point contribue à l'axe

Contribution relative : dans quelle mesure le point est bien représenté par l'axe

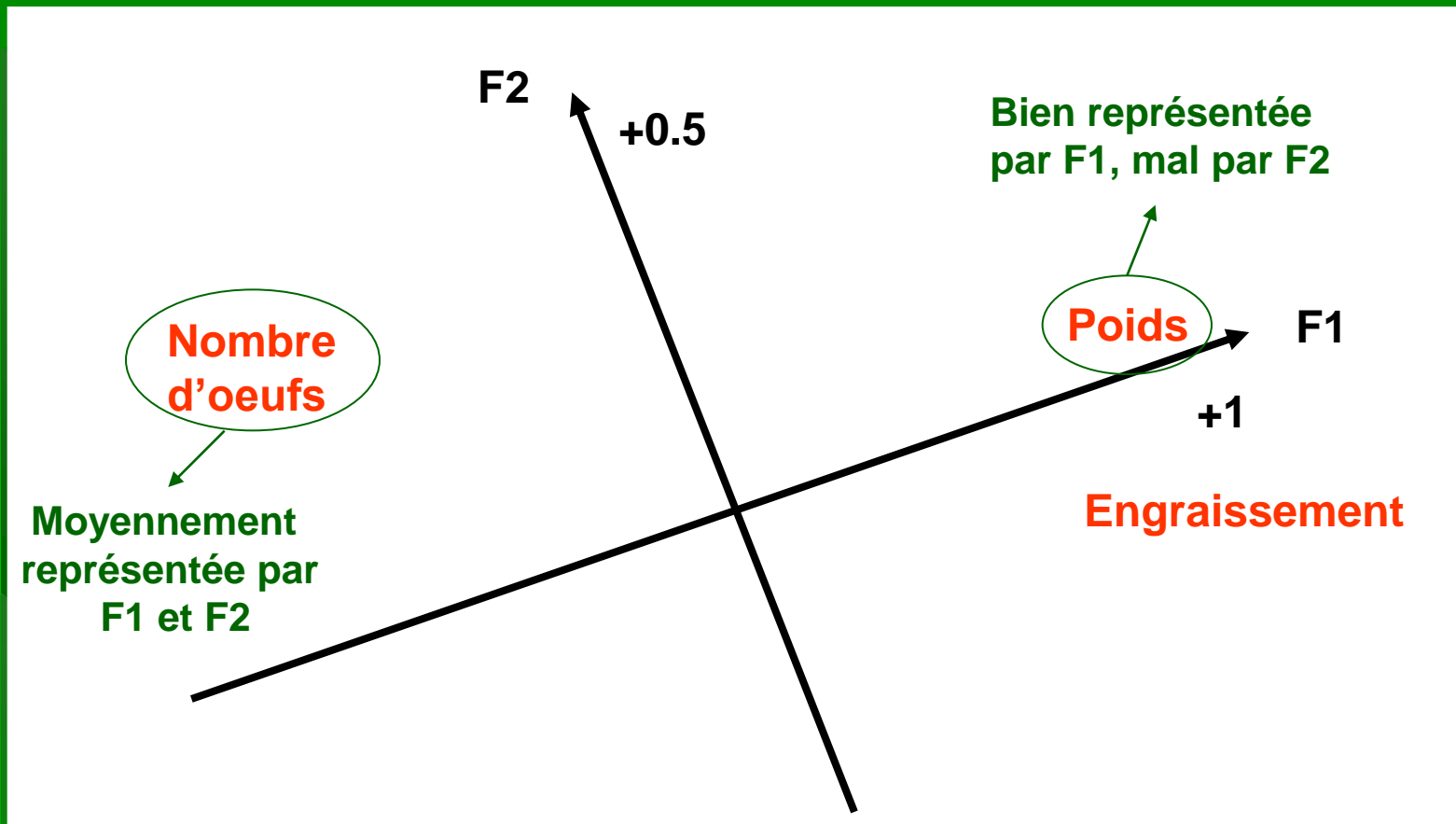
Pour chaque point, la somme des contributions relative de chaque axe vaut 100%

ACP : représentation des variables



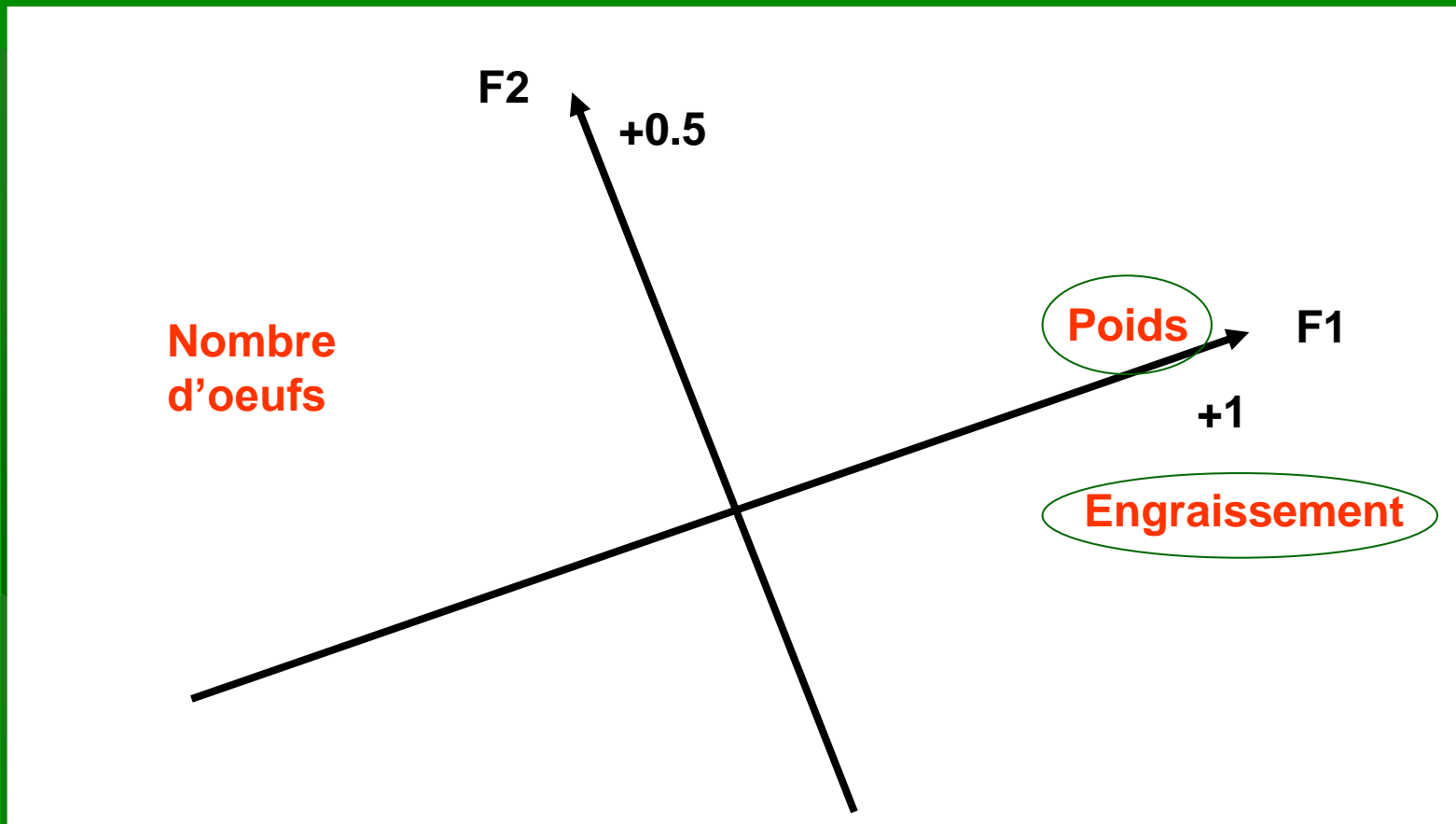
Corrélation entre l'axe et la variable (attention, la corrélation est linéaire, on suppose une distribution normale)

ACP : représentation des variables



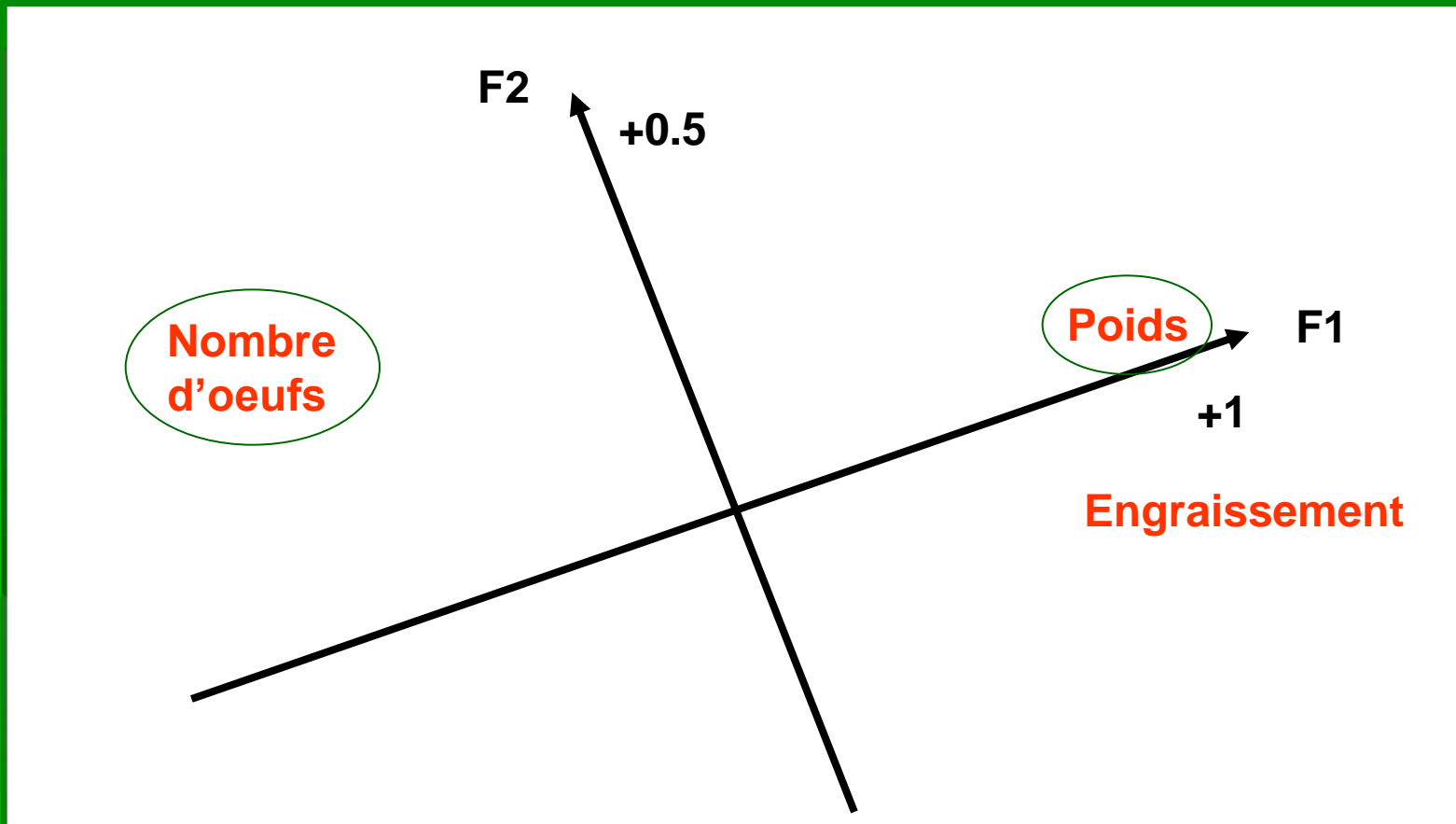
Plus la variable est proche d'un axe, mieux elle est représentée

ACP : interprétation



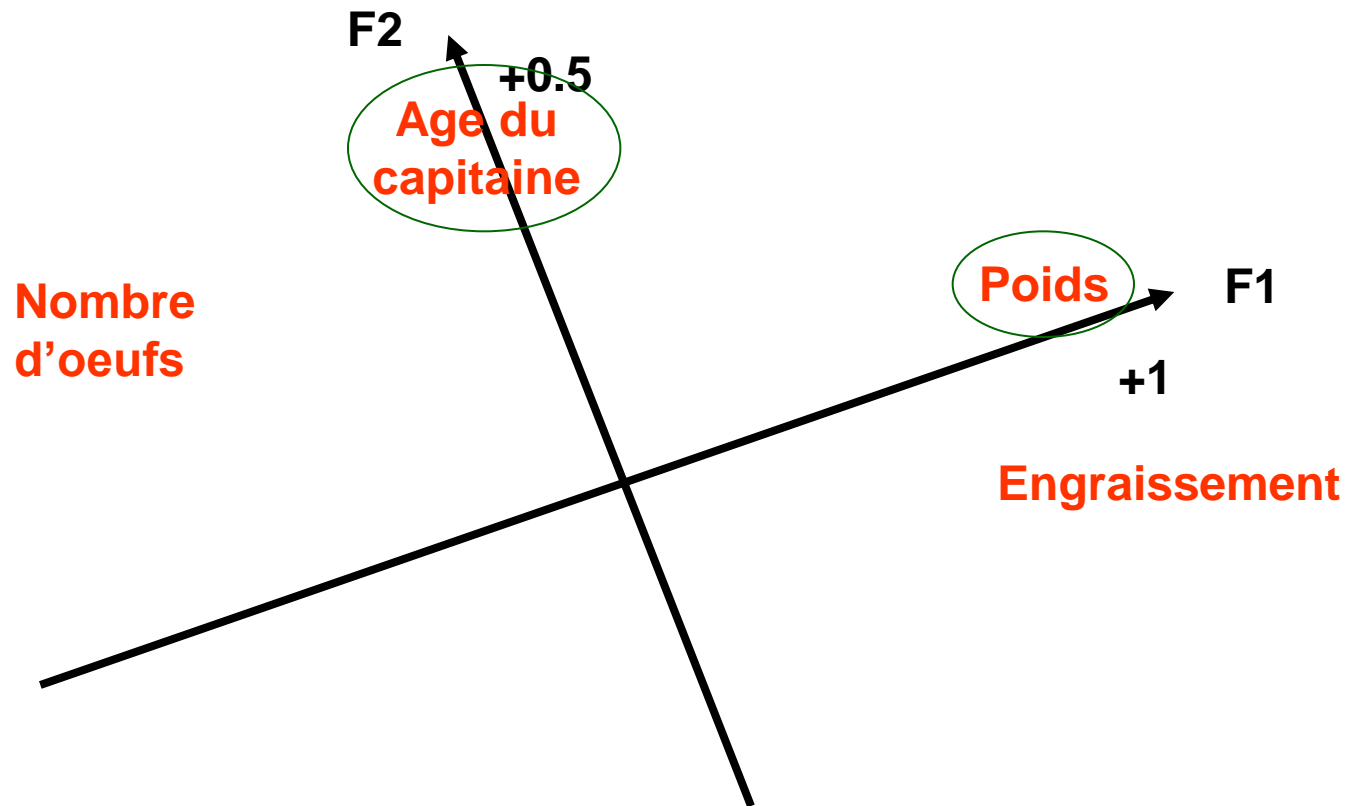
Deux variables corrélées positivement ont des directions similaires

ACP : interprétation



Deux variables corrélées négativement ont des directions opposées

ACP : interprétation



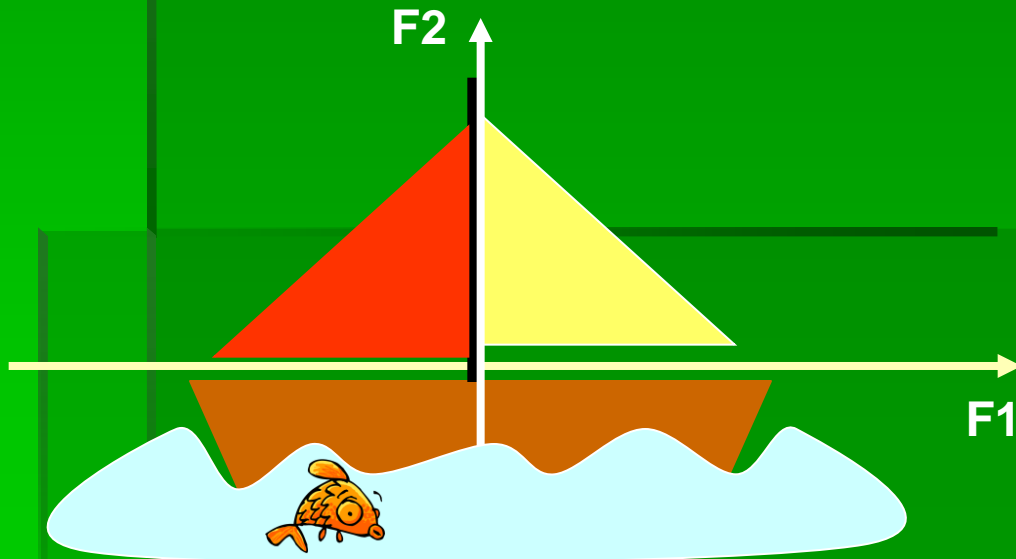
Deux variables non corrélées seront orthogonales

ACP : Résultats

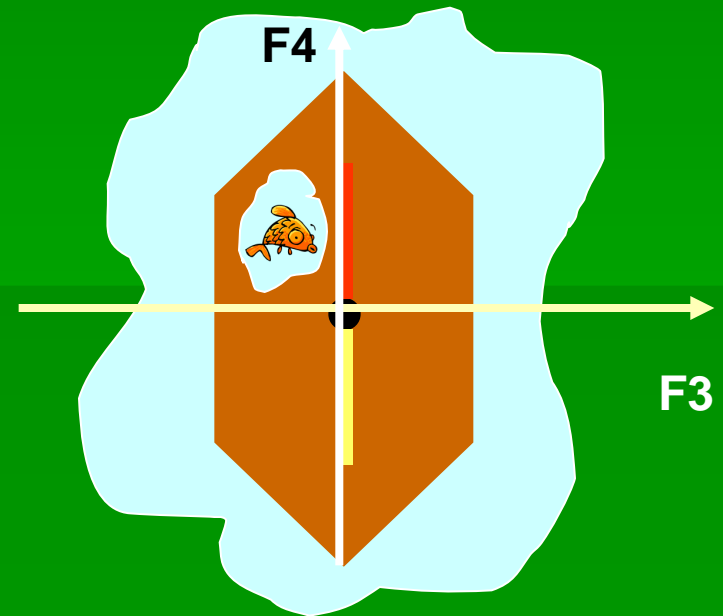
- Facteurs :
 - Valeurs propres
 - Pourcentage de l'inertie totale
- Individus :
 - Contributions
 - Contributions relatives
- Variables :
 - Corrélations entre variables
 - Corrélations variables-facteur

ACP : étapes

- Calcul des axes factoriels
- Projections des du nuage de points sur les plans factoriels ($F1 \times F2$, $F2 \times F5 \dots$)
- Le plan $F1 \times F2$ est la meilleure représentation du nuage de points ... mais pas nécessairement la plus intéressante



Ccl : c'est un bateau !



Ccl : avec le trou qu'il y a au fond, mieux vaut ne pas le prendre !

Logiciels pour l'ACP

- SPAD (clé en main, pratique pour les graphiques)
- SAS (besoin de programmer)
- xlstat
- Hydrolab (clé en main, gratuit, mais pas d'interface graphique)
 - <http://www.cig.ensmp.fr/~hydro/MOD/Hydrolab/hydrolab.htm>

Exemple : la criminalité aux Etats Unis

- Données : nombre de crimes par million de personnes aux Etats-Unis en 1977
 - Vol simple, avec violence, d'automobiles
 - Cambriolage
 - Agression, meurtre, viol

Données

State	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto	Sud-Nord
Alabama	142	252	968	2783	11355	18819	2807	s
Alaska	108	516	968	2840	13317	33698	7533	n
Arizona	95	342	1382	3123	23461	44674	4395	s
Arkansas	88	276	832	2034	9726	18621	1834	s
California	115	494	2870	3580	21394	34998	6635	s
Colorado	63	420	1707	2929	19352	39032	4771	s
Connecticut	42	168	1295	1318	13460	26207	5932	n
Delaware	60	249	1570	1942	16826	36784	4670	n
Florida	102	396	1879	4491	18599	38405	3514	s
Georgia	117	311	1405	2565	13511	21702	2979	s
Hawaii	72	255	1280	641	19115	39204	4894	s
Idaho	55	194	396	1725	10508	25996	2376	n

Sorties d'XLSTAT

Statistiques simples :

Variable	Obs.	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
Murder	50	50	9.000	234.000	78.640	44.100
Rape	50	50	30.000	516.000	251.340	111.746
Robbery	50	50	133.000	4726.000	1273.300	873.382
Assault	50	50	438.000	4853.000	2113.000	1002.530
Burglary	50	50	4461.000	24531.000	12919.040	4324.557
Larceny	50	50	12399.000	44674.000	26712.880	7259.087
Auto	50	50	1444.000	11401.000	3775.260	1933.944

Variable	Modalités	Effectifs	%
Sud-Nord	s	21	42.000
	n	29	58.000

Sorties d'XLSTAT

Analyse en Composantes Principales :

Valeurs propres :

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	3.816	1.221	0.727	0.433	0.384	0.296	0.124
Variabilité (%)	54.516	17.438	10.384	6.180	5.479	4.234	1.769
% cumulé	54.516	71.954	82.338	88.518	93.997	98.231	100.000

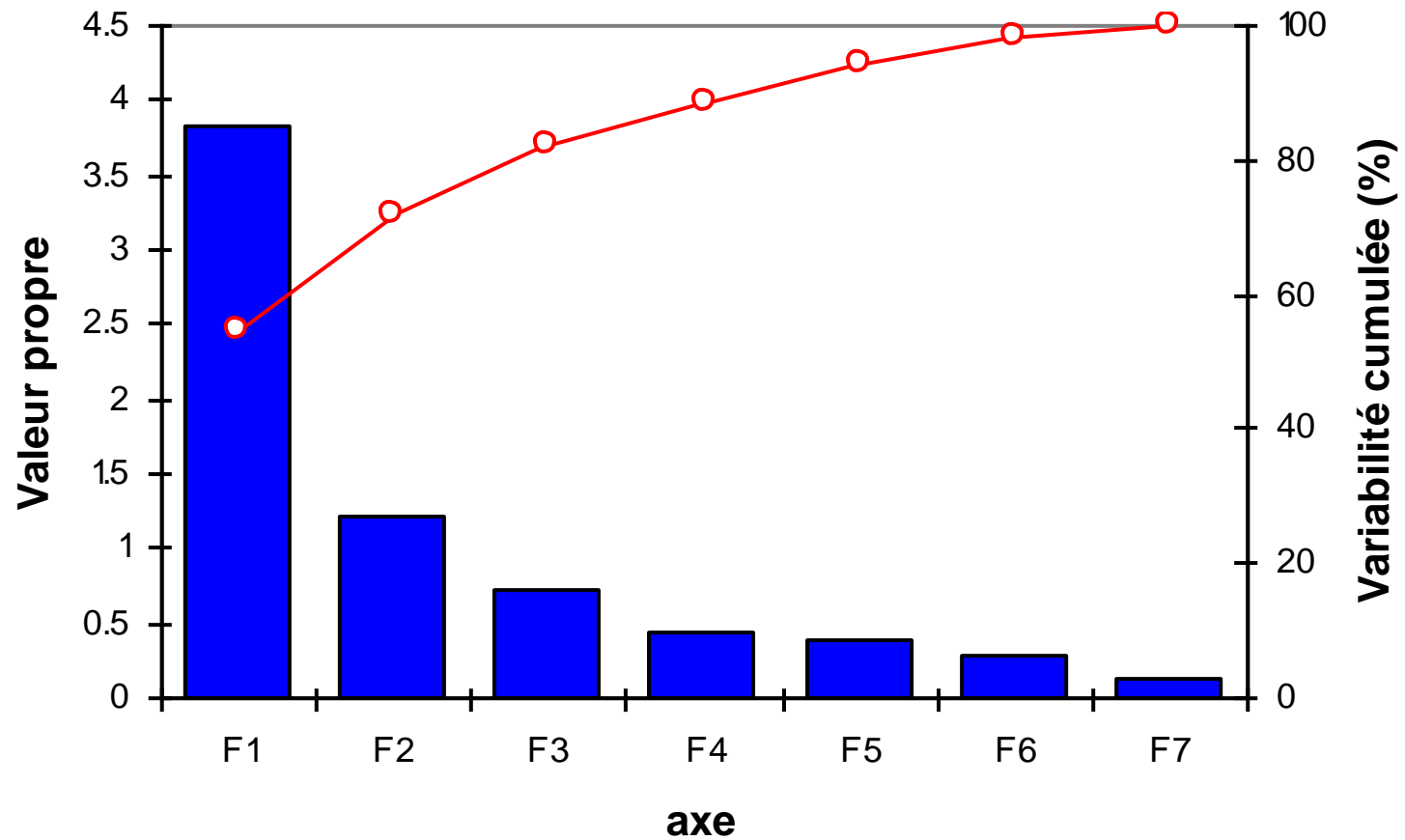
Quels axes faut-il garder ?

❖ Valeurs propres supérieures à 1 : Axes 1 et 2

❖ Variabilité cumulée autour de 80% : Axes 1, 2, 3

Sorties d'XLSTAT

Scree plot



Sorties d'XLSTAT

Corrélation faible \Rightarrow variables indépendantes

Variable bien discriminée sur les axes 1 et 2

Matrice de corrélation (Pearson (n)) :

Variables	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto	Sud	Nord
Murder	1	0.351	0.284	0.529	0.330	0.046	-0.005	0.481	-0.481
Rape	0.351	1	0.549	0.563	0.645	0.616	0.373	0.337	-0.337
Robbery	0.284	0.549	1	0.543	0.582	0.402	0.553	0.089	-0.089
Assault	0.529	0.563	0.543	1	0.623	0.404	0.276	0.469	-0.469
Burglary	0.330	0.645	0.582	0.623	1	0.792	0.558	0.392	-0.392
Larceny	0.046	0.616	0.402	0.404	0.792	1	0.444	0.138	-0.138
Auto	-0.005	0.373	0.553	0.276	0.558	0.444	1	-0.166	0.166
Sud	0.481	0.337	0.089	0.469	0.392	0.138	-0.166	1	-1.000
Nord	-0.481	0.337	-0.089	-0.469	-0.392	-0.138	0.166	-1.000	1

Corrélations élevées \Rightarrow variables corrélées

Variable mal discriminée sur l'axe 3

Sorties d'XLSTAT

Vecteurs propres :

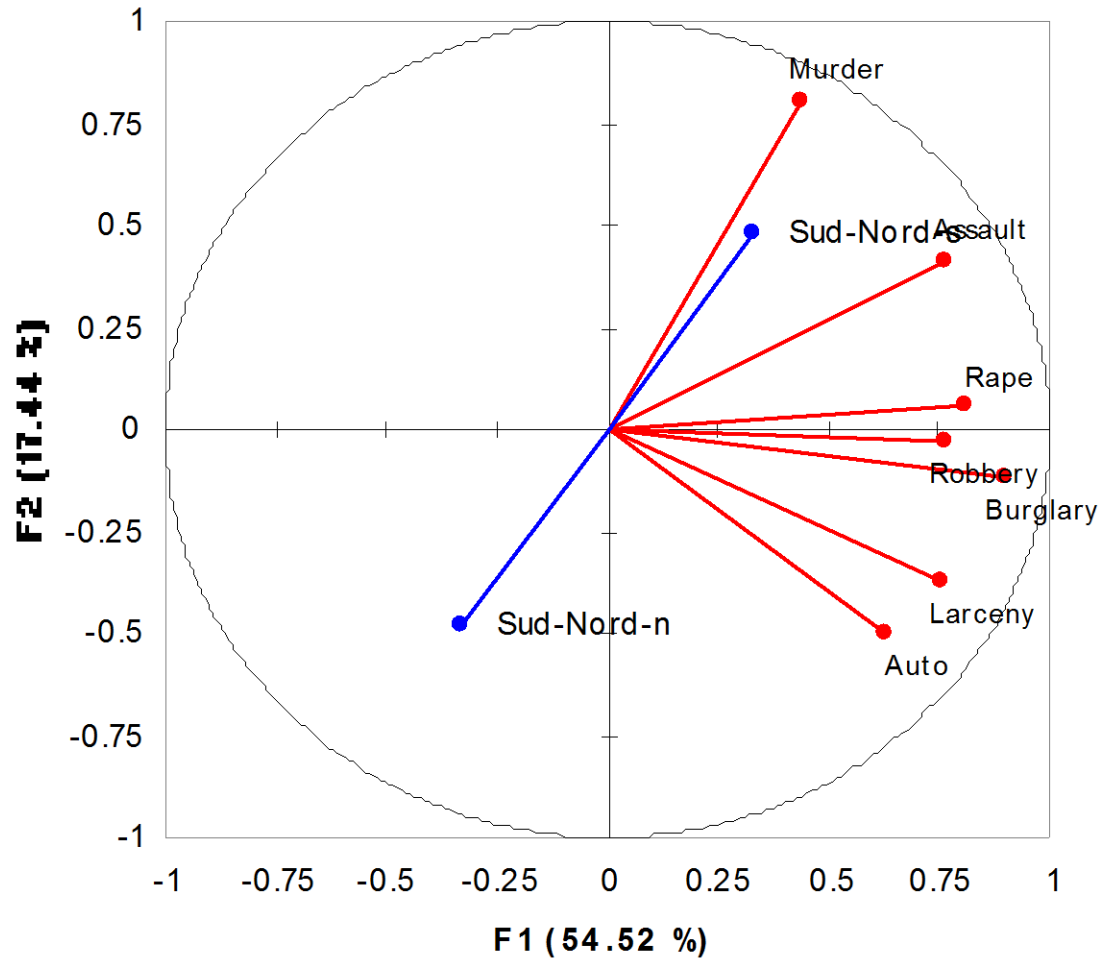
	F1	F2	F3	F4	F5	F6	F7
Murder	0.225	0.731	-0.083	-0.378	-0.435	0.188	-0.202
Rape	0.416	0.055	0.257	0.576	-0.487	-0.410	0.144
Robbery	0.393	-0.023	-0.545	0.475	0.125	0.551	-0.067
Assault	0.391	0.370	-0.015	-0.005	0.700	-0.450	-0.135
Burglary	0.462	-0.103	0.203	-0.352	0.094	0.238	0.739
Larceny	0.388	-0.334	0.547	-0.121	0.033	0.275	-0.590
Auto	0.323	-0.451	-0.539	-0.402	-0.242	-0.399	-0.149

Coordonnées des variables :

	F1	F2	F3	F4	F5	F6	F7
Murder	0.439	0.808	-0.070	-0.249	-0.269	0.102	-0.071
Rape	0.813	0.060	0.219	0.379	-0.301	-0.223	0.051
Robbery	0.768	-0.025	-0.464	0.312	0.077	0.300	-0.024
Assault	0.763	0.409	-0.013	-0.003	0.433	-0.245	-0.048
Burglary	0.903	-0.114	0.173	-0.231	0.058	0.129	0.260
Larceny	0.758	-0.369	0.466	-0.080	0.021	0.150	-0.208
Auto	0.631	-0.498	-0.459	-0.264	-0.150	-0.217	-0.053
Sud-Nord-s	0.332	0.480	0.276	-0.055	0.076	-0.023	0.327
Sud-Nord-n	-0.332	-0.480	-0.276	0.055	-0.076	0.023	-0.327

Sorties d'XLSTAT

Variables (axes F1 et F2 : 71.95 %)



● variables actives ● Variables supplémentaires

Sorties d'XLSTAT

Opposition entre les vols simples et les autres vols

Coordonnées des variables :

	F1	F2	F3	F4	F5	F6	F7
Murder	0.439	0.808	-0.070	-0.249	-0.269	0.102	-0.071
Rape	0.813	0.060	0.219	0.379	-0.301	-0.223	0.051
Robbery	0.768	-0.025	-0.464	0.312	0.077	0.300	-0.024
Assault	0.763	0.409	-0.013	-0.003	0.433	-0.245	-0.048
Burglary	0.903	-0.114	0.173	-0.231	0.058	0.129	0.260
Larceny	0.758	-0.369	0.466	-0.080	0.021	0.150	-0.208
Auto	0.631	-0.498	-0.459	-0.264	-0.150	-0.217	-0.053
Sud-Nord-s	0.332	0.480	0.276	-0.055	0.076	-0.023	0.327
Sud-Nord-n	0.332	-0.480	-0.276	0.055	-0.076	0.023	-0.327

Pas de facteur prédominant :
L'axe 1 rend compte de la criminalité générale

Opposition entre les crimes envers les personnes et ceux aux biens

Sorties d'XLSTAT

Contributions des variables
(%) :

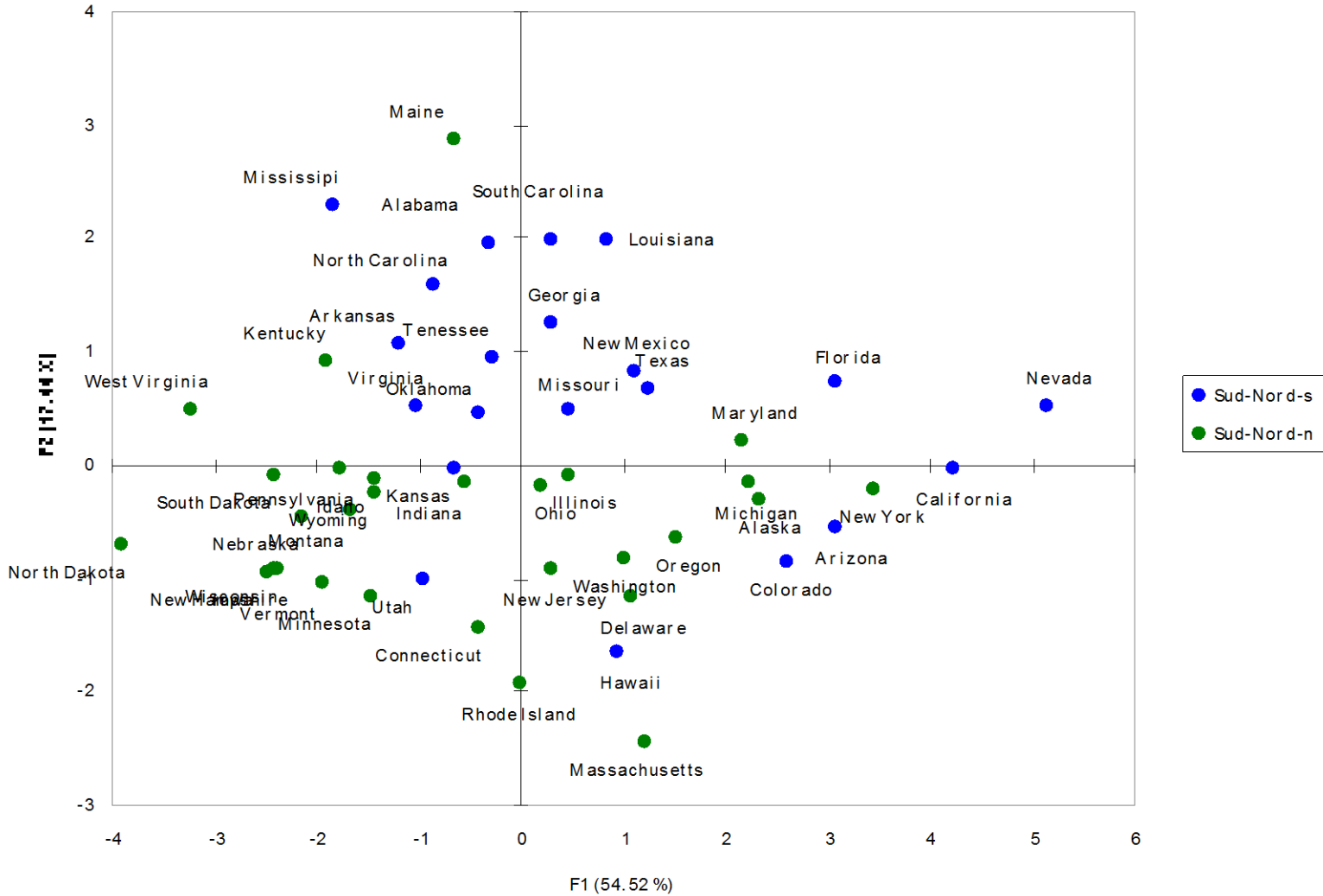
	F1	F2	F3	F4	F5	F6	F7
Murder	5.053	53.436	0.682	14.315	18.912	3.525	4.077
Rape	17.336	0.299	6.605	33.179	23.696	16.798	2.087
Robbery	15.462	0.051	29.649	22.528	1.551	30.309	0.449
Assault	15.275	13.672	0.024	0.003	48.977	20.214	1.835
Burglary	21.385	1.056	4.103	12.376	0.877	5.654	54.549
Larceny	15.054	11.141	29.897	1.463	0.111	7.561	34.774
Auto	10.435	20.345	29.040	16.136	5.876	15.938	2.229

Sorties d'XLSTAT

Coordonnées des observations :

Observation	F1	F2	F3	F4	F5	F6	F7
Alabama	-0.305	1.951	-0.339	-0.252	-0.154	-0.415	0.093
Alaska	2.349	-0.318	0.233	0.013	-1.423	-1.872	-0.663
Arizona	3.086	-0.541	1.786	-0.914	0.401	0.490	0.193
Arkansas	-1.200	1.061	0.098	0.611	-0.183	-0.378	0.298
California	4.250	-0.028	-0.307	0.381	-0.304	-0.199	0.384
Colorado	2.601	-0.847	1.097	0.303	0.123	-0.165	0.169
Connecticut	-0.411	-1.449	-0.746	-0.589	-0.089	0.086	0.134
Delaware	1.080	-1.157	0.545	-0.365	0.137	0.611	-0.139
Florida	3.077	0.715	1.105	0.262	1.106	-0.310	-0.250
Georgia	0.319	1.243	-0.153	0.251	-0.216	-0.170	0.404
Hawaii	0.935	-1.650	0.968	-0.866	-0.935	1.218	0.193
Idaho	-1.425	-0.125	0.695	-0.069	0.209	-0.142	-0.093
			...				
Sud	0.761	0.623	0.277	-0.042	0.055	-0.015	0.135
Nord	-0.551	-0.451	-0.200	0.031	-0.040	0.011	-0.098

Observations (axes F1 et F2 : 71.95 %)

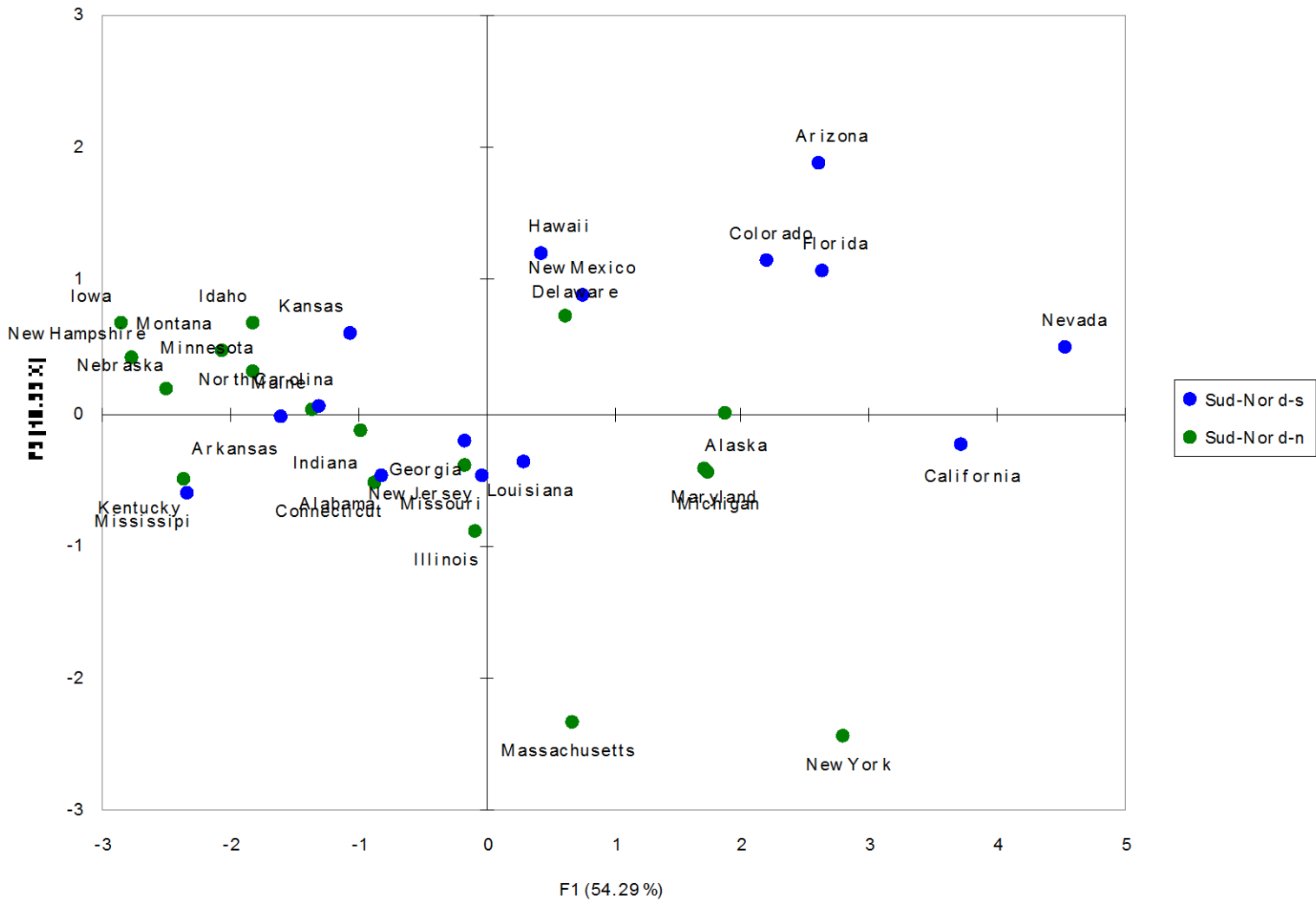


Sorties d'XLSTAT

Contributions des observations (%)

	F1	F2	F3	F4	F5	F6	F7
Alabama	0.049	6.238	0.315	0.294	0.123	1.162	0.141
Alaska	2.892	0.166	0.149	0.001	10.560	23.656	7.106
Arizona	4.990	0.480	8.781	3.862	0.840	1.618	0.601
Arkansas	0.755	1.844	0.027	1.728	0.174	0.967	1.438
California	9.465	0.001	0.259	0.672	0.481	0.268	2.382
Colorado	3.545	1.175	3.311	0.423	0.079	0.183	0.461
Connecticut	0.088	3.441	1.529	1.603	0.042	0.050	0.290
Delaware	0.611	2.194	0.817	0.616	0.098	2.517	0.311
Florida	4.963	0.838	3.361	0.317	6.384	0.648	1.008
Georgia	0.053	2.531	0.064	0.291	0.244	0.195	2.636
Hawaii	0.459	4.463	2.581	3.468	4.560	10.009	0.599
Idaho	1.064	0.026	1.330	0.022	0.228	0.135	0.139
Illinois	0.115	0.012	3.022	0.018	0.170	1.007	10.424
Indiana	0.150	0.040	0.064	0.389	1.203	0.002	0.136
Iowa	3.224	1.490	0.852	0.002	0.496	0.918	2.787
Kansas	0.212	0.004	0.803	0.035	0.160	0.707	0.134

Observations (axes F1 et F3 : 65.22 %)



Massachusetts	0.772	9.767	18.364	8.073	0.019	14.479	0.774
Michigan	2.652	0.046	0.854	2.288	0.321	0.050	0.714
Minnesota	1.100	2.275	0.094	0.109	0.166	0.366	0.755
Mississippi	1.731	8.578	0.607	0.096	1.286	0.044	2.786
Missouri	0.117	0.376	0.770	0.632	0.028	0.190	0.275
Montana	1.420	0.267	0.566	0.150	0.042	0.261	8.467
Nebraska	2.367	0.338	0.071	0.758	0.014	0.004	1.021
Nevada	13.954	0.424	0.361	0.005	1.076	5.525	0.614
New Hampshire	2.966	1.383	0.246	1.042	0.005	0.257	0.615
New Jersey	0.048	1.354	1.037	0.095	0.143	0.838	0.176
New Mexico	0.673	1.084	3.014	1.826	0.797	3.725	0.001
New York	6.251	0.077	23.010	2.373	0.927	9.567	0.216

Application : Logiciel SPAD

- Base de données : Iris
- Variables mesurées
 - SELO - Longueur du sépale
 - SELA - Largeur du sépale
 - PELO - Longueur du pétale
 - PELA - Largeur du pétale
- 150 Individus
- 3 espèces : Setosa, Versicolor, Virginica



Application : Logiciel SPAD

ANALYSE EN COMPOSANTES PRINCIPALES

STATISTIQUES SOMMAIRES DES VARIABLES CONTINUES

EFFECTIF TOTAL : 150 POIDS TOTAL : 150.00

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
1 . SELO - Longueur du sépale	150	150.00	58.43	8.25	43.00	79.00
2 . SELA - Largeur du sépale	150	150.00	30.57	4.34	20.00	44.00
3 . PELO - Longueur du pétale	150	150.00	37.58	17.59	10.00	69.00
4 . PELA - Largeur du pétale	150	150.00	11.99	7.60	1.00	25.00

MATRICE DES CORRELATIONS

	SELO	SELA	PELO	PELA
SELO	1.00			
SELA	-0.12	1.00		
PELO	0.87	-0.43	1.00	
PELA	0.82	-0.37	0.96	1.00

Application : Logiciel SPAD

MATRICE DES VALEURS-TESTS

| SELO SELA PELO PELA

```
-----+-----
SELO | 99.99
SELA | -1.45  99.99
PELO | 16.42  -5.61  99.99
PELA | 14.09  -4.70  24.30  99.99
-----+-----
```

VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 4.0000

SOMME DES VALEURS PROPRES 4.0000

HISTOGRAMME DES 4 PREMIERES VALEURS PROPRES

```
+-----+-----+-----+-----+-----+-----+
| NUMERO | VALEUR | POURCENTAGE | POURCENTAGE |
|        | PROPRE |              | CUMULE      |
+-----+-----+-----+-----+-----+-----+
|  1    | 2.9185 | 72.96       | 72.96       | *****
|  2    | 0.9140 | 22.85       | 95.81       | *****
|  3    | 0.1468 | 3.67        | 99.48       | *****
|  4    | 0.0207 | 0.52        | 100.00      | *
+-----+-----+-----+-----+-----+-----+
```

Les VP intéressantes sont celles >1

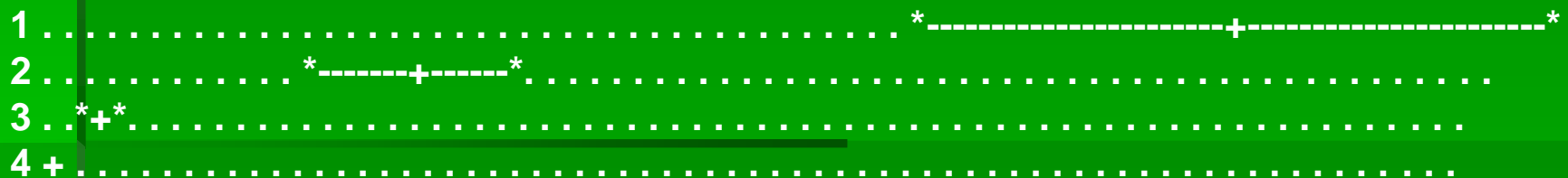
Application : Logiciel SPAD

INTERVALLES LAPLACIENS D'ANDERSON

INTERVALLES AU SEUIL 0.95

NUMERO	BORNE INFERIEURE	VALEUR PROPRE	BORNE SUPERIEURE
1	2.2558	2.9185	3.5812
2	0.7065	0.9140	1.1216
3	0.1134	0.1468	0.1801
4	0.0160	0.0207	0.0254

ETENDUE ET POSITION RELATIVE DES INTERVALLES



Moins les intervalles se chevauchent d'un axe à l'autre, plus l'axe apporte une information différente

Application : Logiciel SPAD

COORDONNEES DES VARIABLES SUR LES AXES 1 A 4 VARIABLES ACTIVES

VARIABLES IDEN - LIBELLE COURT	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
SELO - Longueur du sépale	0.89	-0.36	-0.28	-0.04	0.00	0.89	-0.36	-0.28	-0.04	0.00	0.52	-0.38	-0.72	-0.26	0.00
SELA - Largeur du sépale	-0.46	-0.88	0.09	0.02	0.00	-0.46	-0.88	0.09	0.02	0.00	-0.27	-0.92	0.24	0.12	0.00
PELO - Longueur du pétale	0.99	-0.02	0.05	0.12	0.00	0.99	-0.02	0.05	0.12	0.00	0.58	-0.02	0.14	0.80	0.00
PELA - Largeur du pétale	0.96	0.06	0.24	-0.08	0.00	0.96	-0.06	0.24	-0.08	0.00	0.56	-0.07	0.63	-0.52	0.00

L'axe 1 est défini par les variables SELO, SELA, PELA, mais pas par PELO

L'axe 2 est défini par la variable PELO mais pas par SELO, SELA, PELA

Application : Logiciel SPAD

COORDONNEES ET VALEURS-TEST DES MODALITES AXES 1 A 4

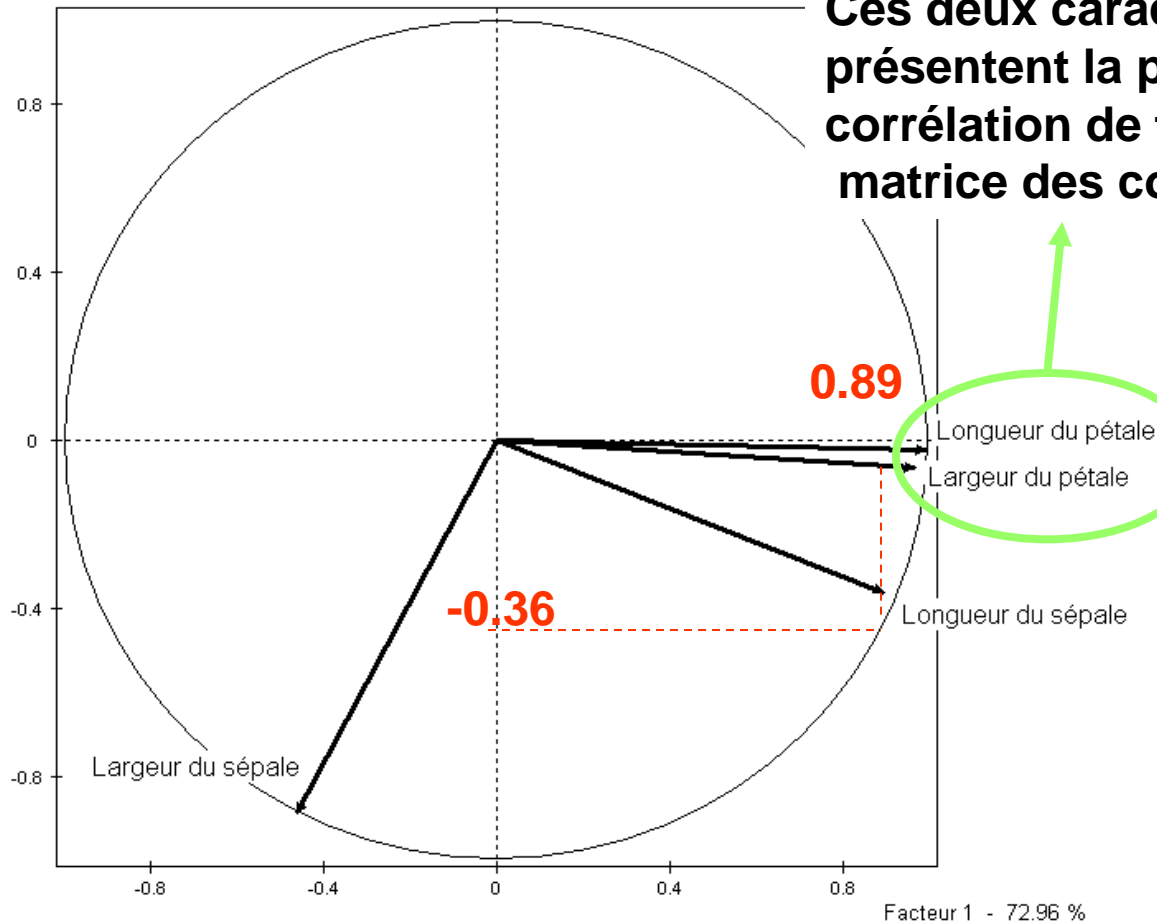
MODALITES			VALEURS-TEST					COORDONNEES					
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	0	1	2	3	4	0	DISTO.
5 . Les 3 espèces													
AA_1 - SETOSA	50	50.00	-11.2	-2.6	-1.0	-1.1	0.0	-2.22	-0.29	-0.04	-0.02	0.00	5.04
AA_2 - VERSICOLOR	50	50.00	2.5	5.0	-2.2	1.8	0.0	0.50	0.55	-0.10	0.03	0.00	0.56
AA_3 - VIRGINICA	50	50.00	8.7	-2.4	3.1	-0.7	0.0	1.73	-0.26	0.14	-0.01	0.00	3.07

L'axe 1 oppose
SETOSA
à
VERSICOLOR ET VIRGINICA

L'axe 2 oppose
VERSICOLOR
à
SETOSA ET VIRGINICA

REPRESENTATION GRAPHIQUE : CERCLE DES CORRELATIONS AXES 1 ET 2

Facteur 2 - 22.85 %



Ces deux caractères
présentent la plus forte
corrélation de toute la
matrice des corrélations

0.89

Longueur du pétale

Largeur du pétale

Longueur du sépale

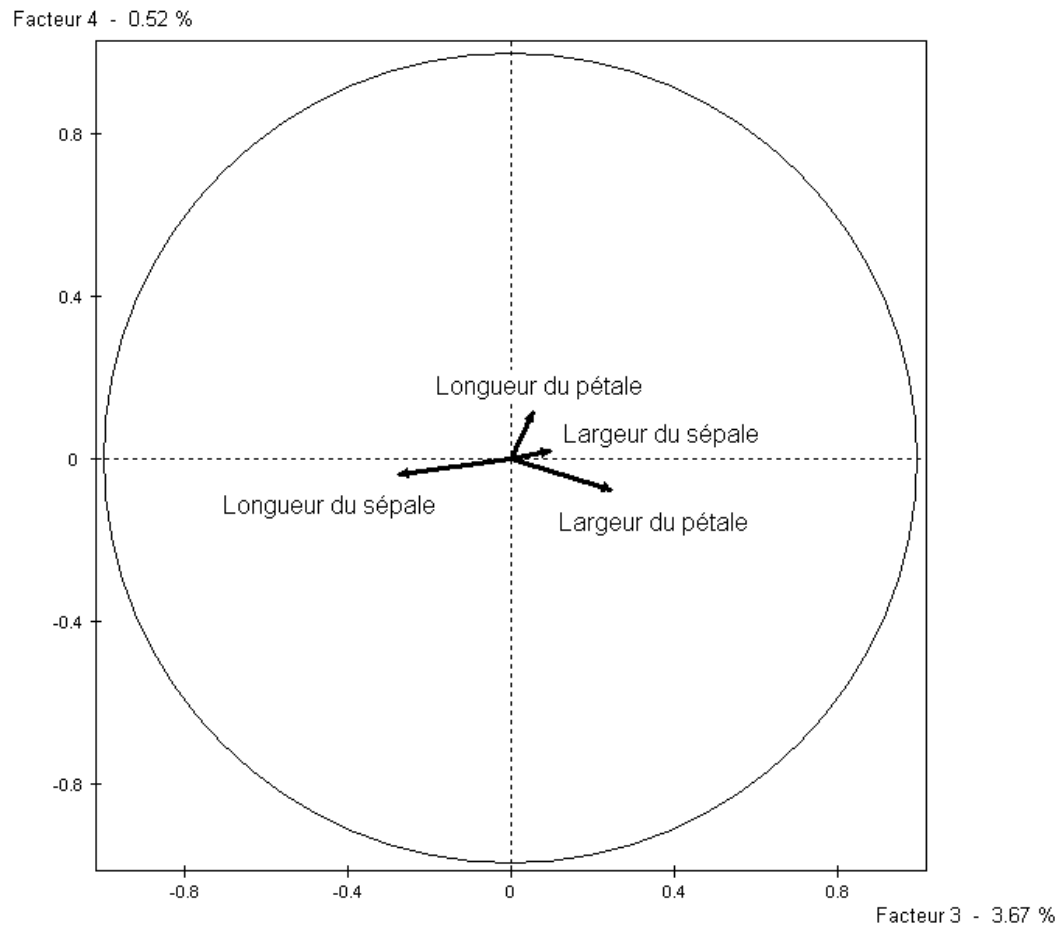
-0.36

Largeur du sépale

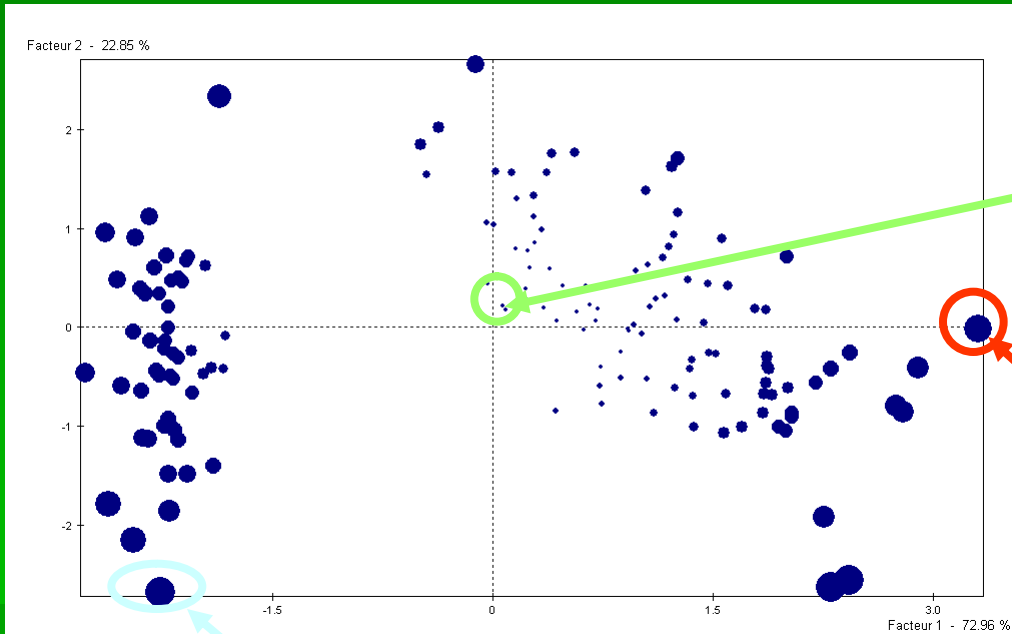
Facteur 1 - 72.96 %

L'axe 1 représente la plus grosse corrélation négative (ici, entre SELA ET PELO), même si elle n'est pas très élevée

REPRESENTATION GRAPHIQUE : CERCLE DES CORRELATIONS AXES 3 ET 4



REPRESENTATION GRAPHIQUE : INDIVIDUS SUR LES AXES 1 ET 2

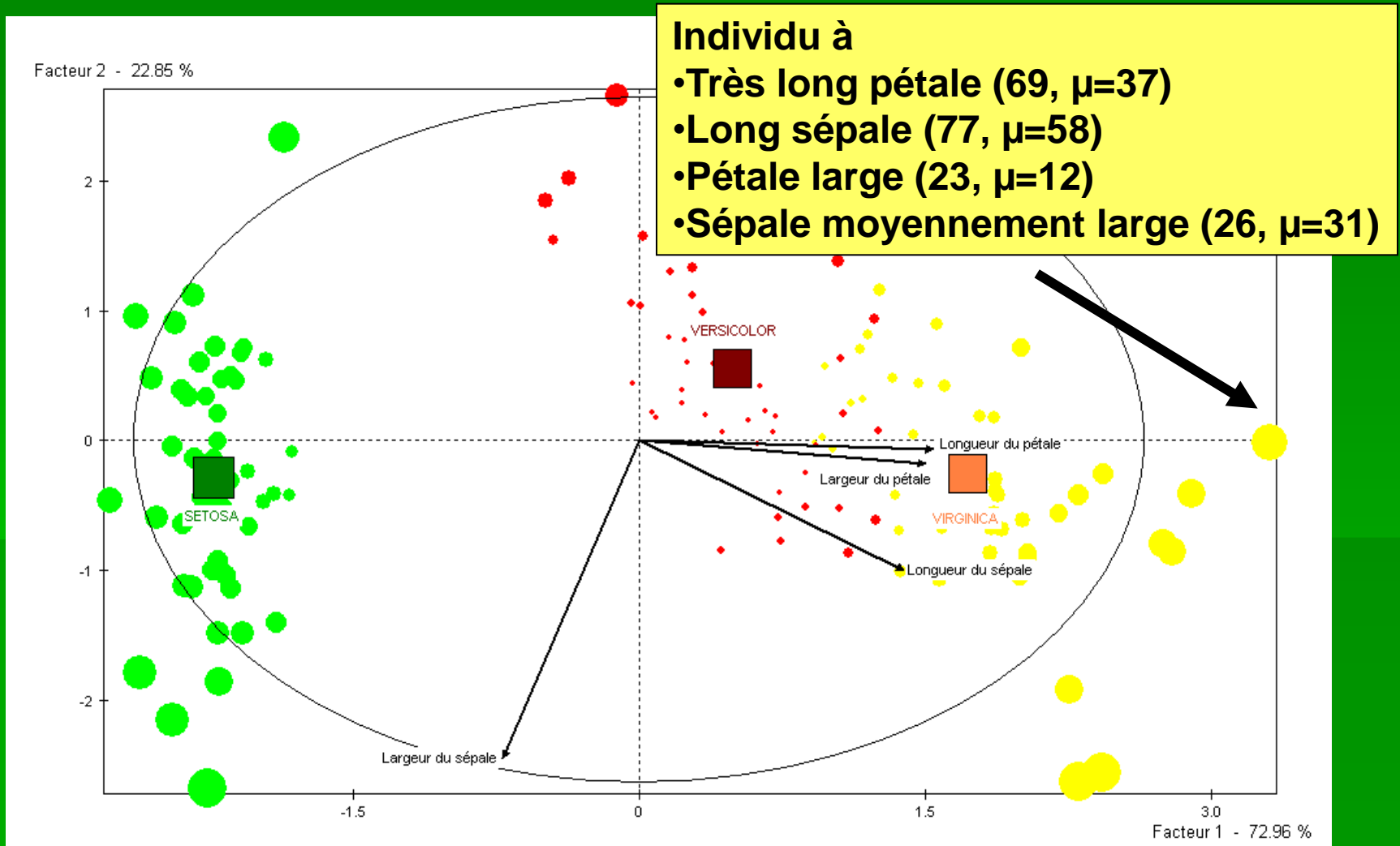


Individu à faible
contribution sur l'axe 1
(mal représenté par cet axe)

Individu à forte
contribution sur l'axe 1

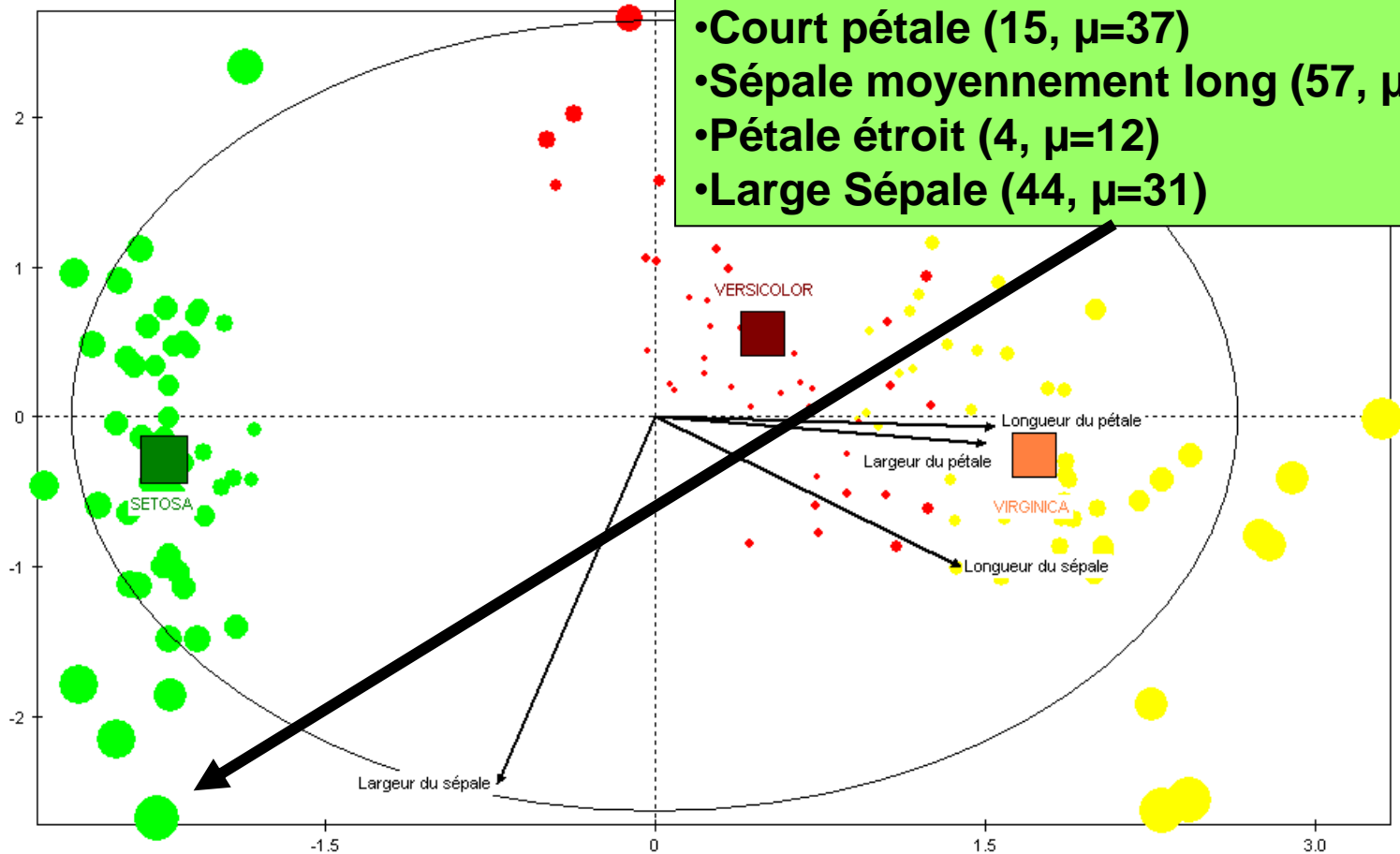
Individu à forte
contribution sur l'axe 2

REPRESENTATION GRAPHIQUE : INDIVIDUS SUR LES AXES 1 ET 2



REPRESENTATION GRAPHIQUE : INDIVIDUS SUR LES AXES 1 ET 2

Facteur 2 - 22.85 %

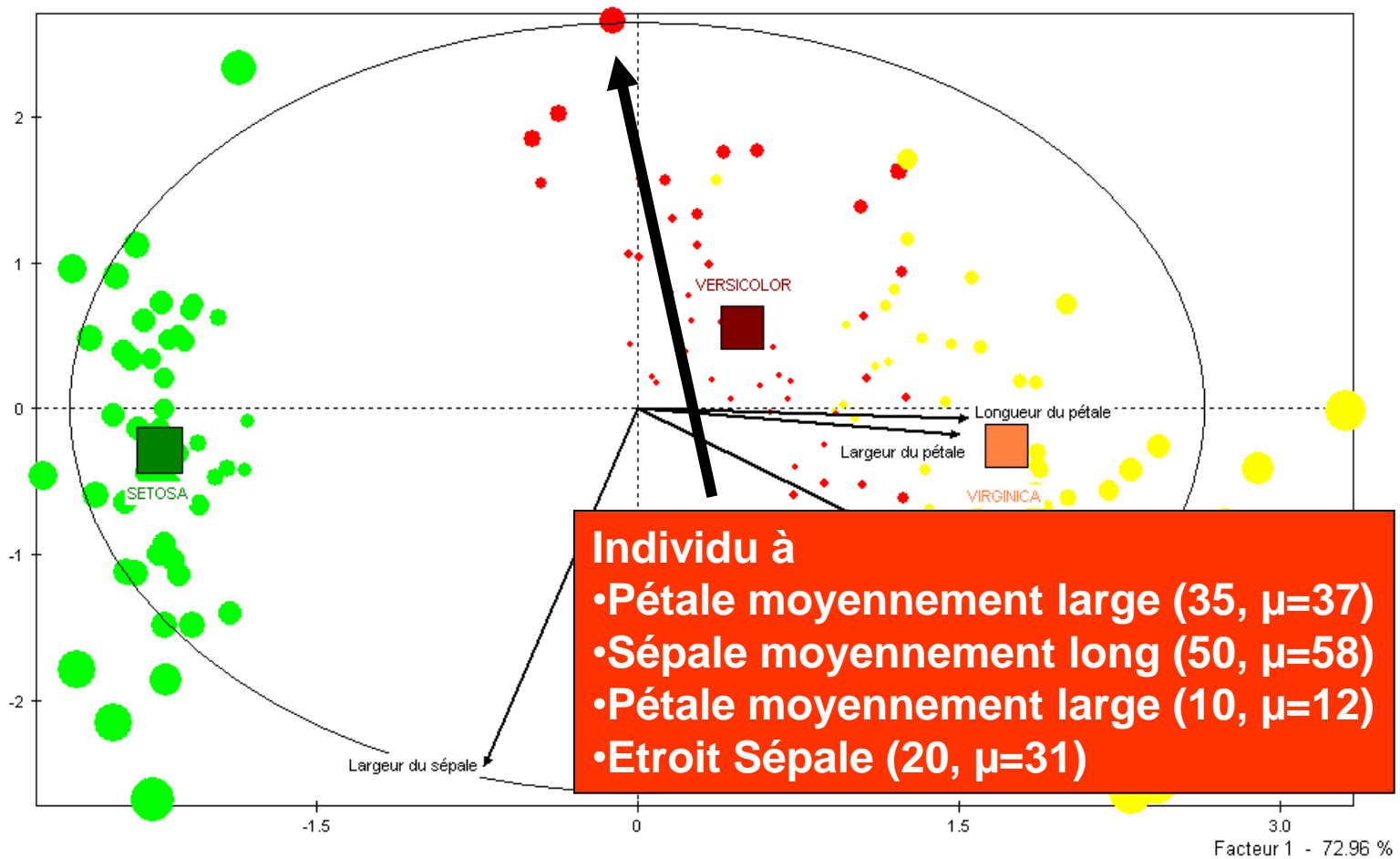


Individu à

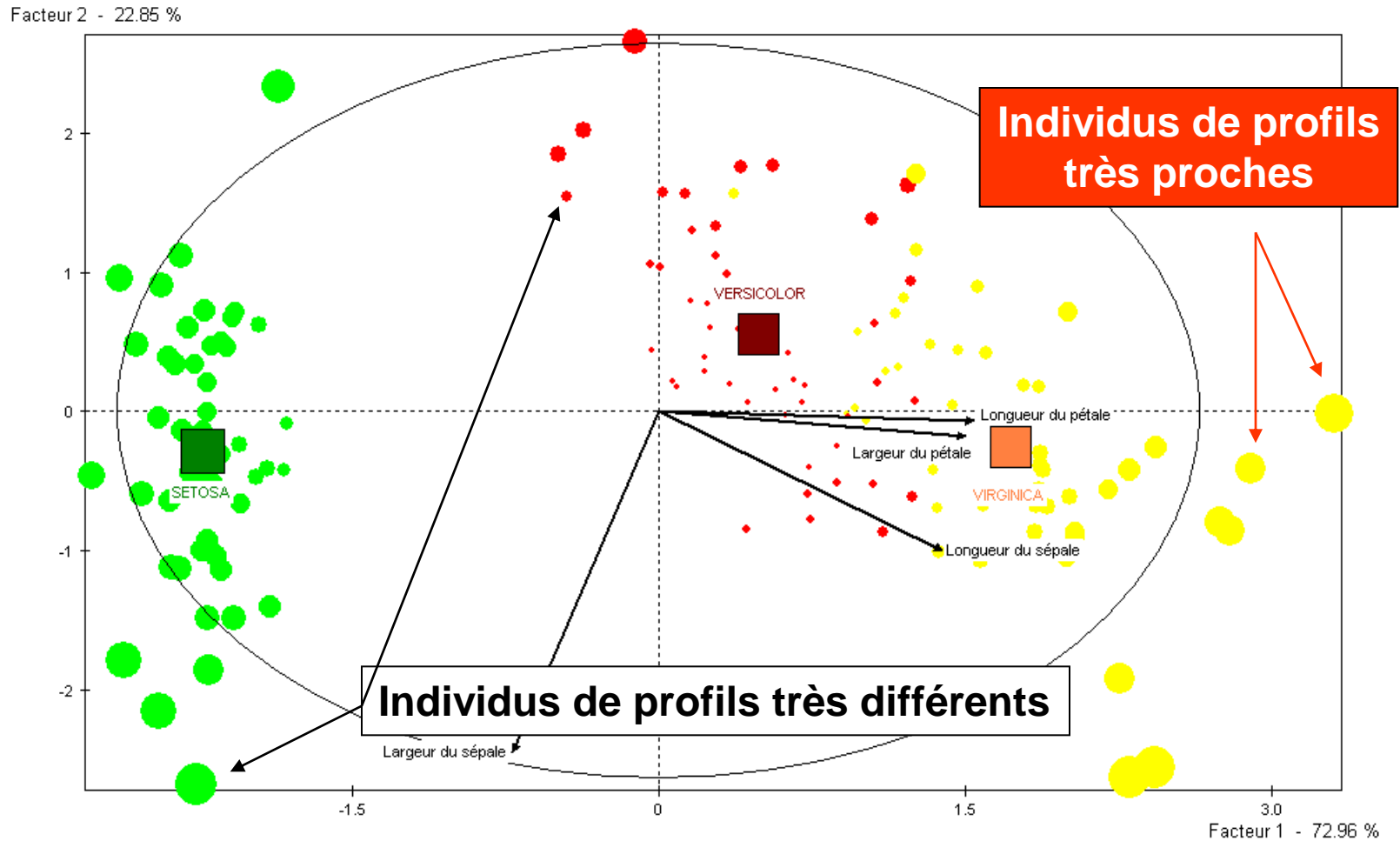
- Court pétale (15, $\mu=37$)
- Sépale moyennement long (57, $\mu=58$)
- Pétale étroit (4, $\mu=12$)
- Large Sépale (44, $\mu=31$)

REPRESENTATION GRAPHIQUE : INDIVIDUS SUR LES AXES 1 ET 2

Facteur 2 - 22.85 %



REPRESENTATION GRAPHIQUE : INDIVIDUS SUR LES AXES 1 ET 2



ACP : ANALYSE NORMEE OU NON NORMEE ?

- Certaines variables ont une forte variance, d'autres beaucoup moins
- L'analyse non normée donne à chaque variable un poids proportionnel à leur variance
 - **Cas des tableaux de notes ou de rang**
- L'analyse normée utilise les valeurs réduites, toutes les variables ont le même poids
 - **Tous les autres cas, impérativement en cas de variances hétérogènes**

ACP : effet d'un individu aberrant

Corrélations entre variables

- avec les données originales
- avec un individu aberrant

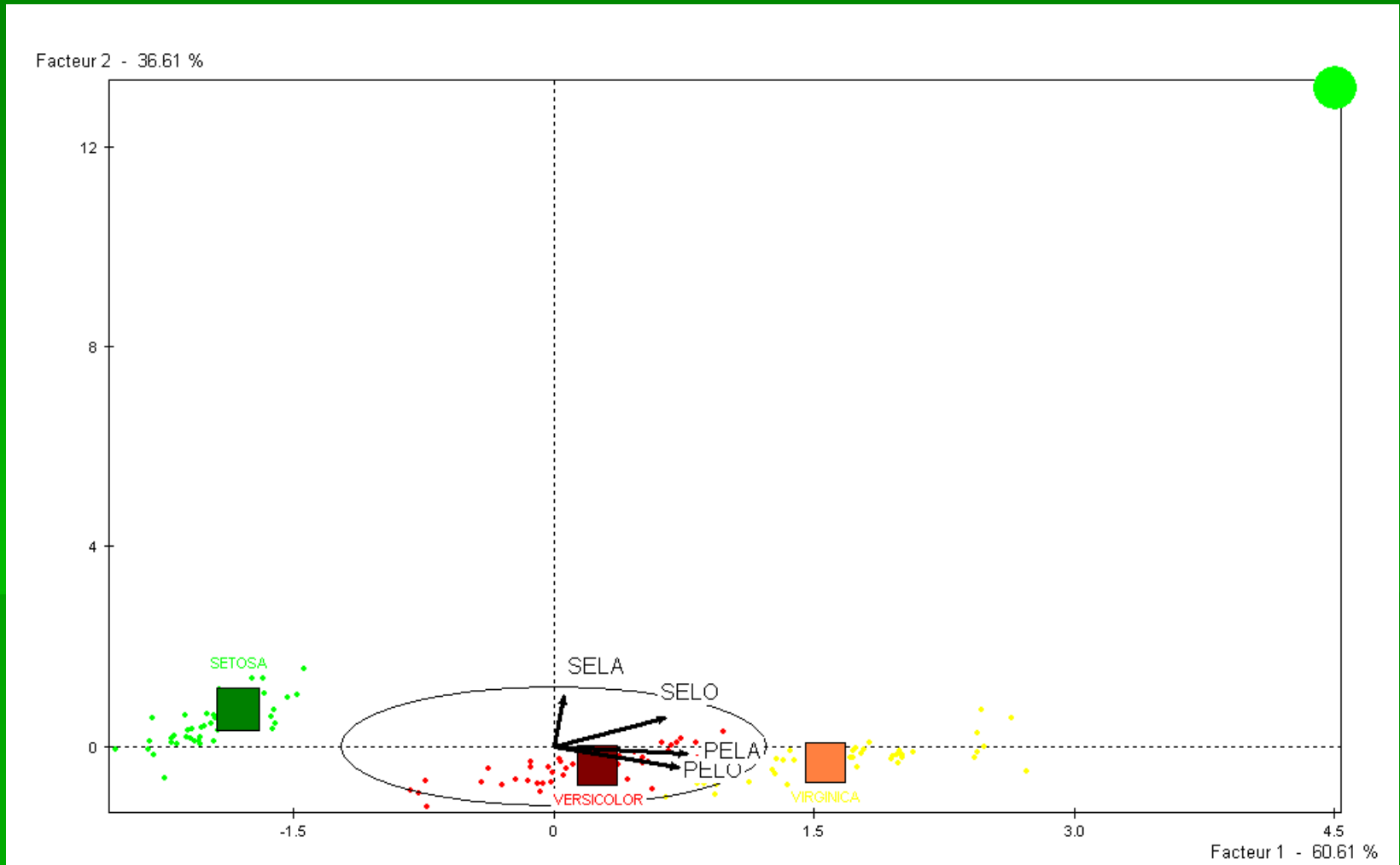


	SELO	SELA	PELO	PELA
SELO	1.00	-0.12	0.87	0.82
SELA	0.57	1.00	-0.43	-0.37
PELO	0.52	-0.33	1.00	0.96
PELA	0.67	-0.06	0.93	1.00

Valeurs propres des facteurs 1 et 2 et % d'inertie

	V.P.	V.P.	INERTIE	INERTIE
	F1	F2	F1	F2
Données originales	2.92	0.91	72.96%	22.85%
Individu aberrant	2.42	1.46	60.61%	36.61%

ACP : effet d'un individu aberrant



Exercice : les prix et l'euro

- Ouvrir le fichier « prix.xls »
- Réaliser les statistiques élémentaires sur les variables
- Réaliser les tableaux croisés dynamiques pour la variables « euro »
- Réaliser l'ACP

Statistiques élémentaires

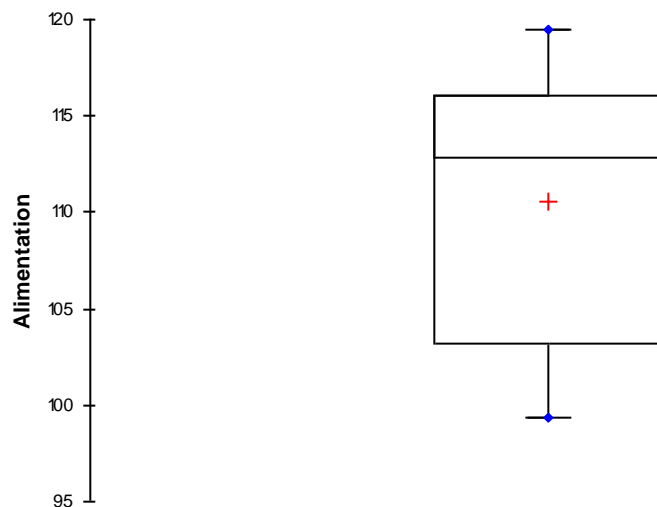
- Menu XLSTAT
- Description des données
- ⇒ Statistiques descriptives

Statistique	Alimentation	Pain	Viande	Loyers	Carburants
Nb. d'observations	39	39	39	39	39
Nb. de valeurs manquantes	0	0	0	0	0
Somme des poids	39	39	39	39	39
Minimum	99.400	99.800	98.700	99.025	93.600
Maximum	119.520	116.490	123.580	113.680	142.260
Eff. du minimum	1	1	1	1	1
Eff. du maximum	1	1	1	1	1
Amplitude	20.120	16.690	24.880	14.655	48.660
1er Quartile	103.200	103.200	102.500	102.988	108.130
Médiane	112.800	110.800	112.700	105.900	112.320
3ème Quartile	116.050	114.300	117.500	109.860	122.663
Somme	4311.470	4253.920	4343.940	4149.550	4478.038
Moyenne	110.551	109.075	111.383	106.399	114.821

Statistique	Alimentation	Pain	Viande	Loyers	Carburants
Variance (n)	44.589	32.495	62.872	18.379	151.140
Variance (n-1)	45.763	33.350	64.526	18.863	155.117
Ecart-type (n)	6.678	5.700	7.929	4.287	12.294
Ecart-type (n-1)	6.765	5.775	8.033	4.343	12.455
Coefficient de variation	0.060	0.052	0.071	0.040	0.107
Asymétrie (Pearson)	-0.490	-0.388	-0.363	0.018	0.276
Asymétrie (Fisher)	-0.510	-0.404	-0.378	0.019	0.287
Asymétrie (Bowley)	-0.494	-0.369	-0.360	0.152	0.423
Aplatissement (Pearson)	-1.319	-1.454	-1.251	-1.059	-0.378
Aplatissement (Fisher)	-1.334	-1.488	-1.257	-1.037	-0.260
Ecart-type de la moyenne	1.083	0.925	1.286	0.695	1.994
Borne inf. de la moyenne (95%)	108.358	107.203	108.779	104.991	110.784
Borne sup. de la moyenne (95%)	112.743	110.947	113.987	107.807	118.859

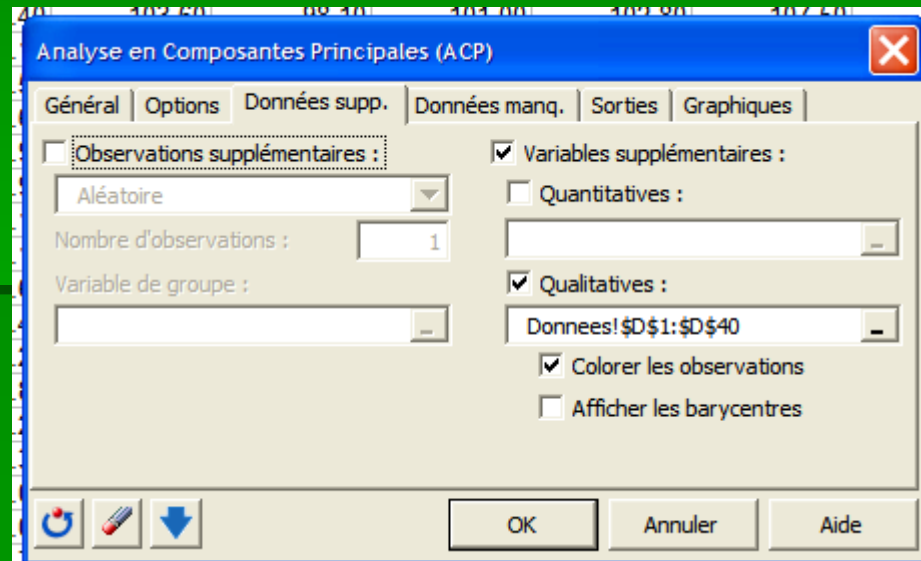
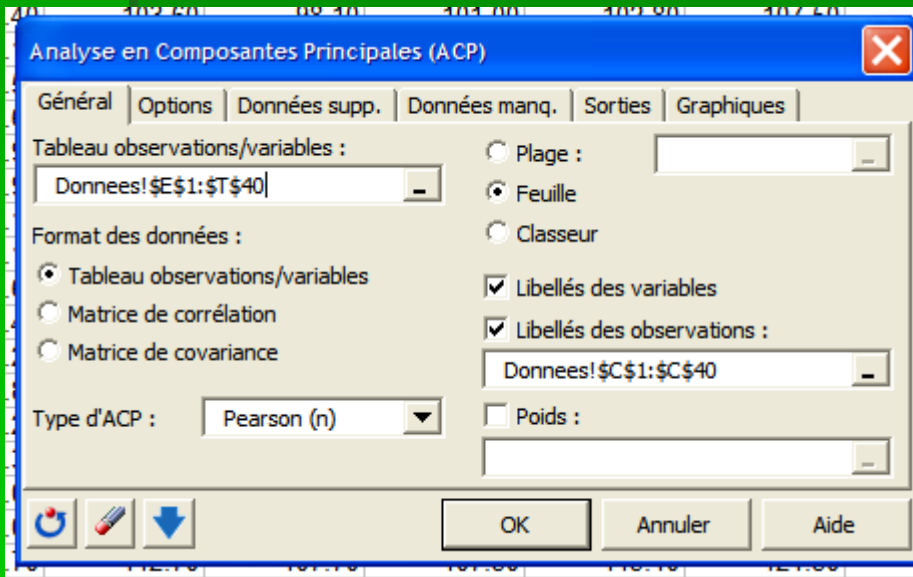
Statistique	Alimentation	Pain	Viande	Loyers	Carburants
Ecart-type(Asymétrie (Fisher))	0.378	0.378	0.378	0.378	0.378
Ecart-type(Aplatissement (Fisher))	0.741	0.741	0.741	0.741	0.741
Ecart absolu moyen	5.895	5.168	6.804	3.592	9.708
Ecart absolu médian	4.300	4.000	5.700	3.445	6.620
Moyenne géométrique	110.344	108.924	111.095	106.312	114.169
Ecart-type géométrique	1.064	1.055	1.076	1.042	1.114
Moyenne harmonique	110.134	108.770	110.802	106.226	113.521

Box plot (Alimentation)



Réalisation de l'ACP

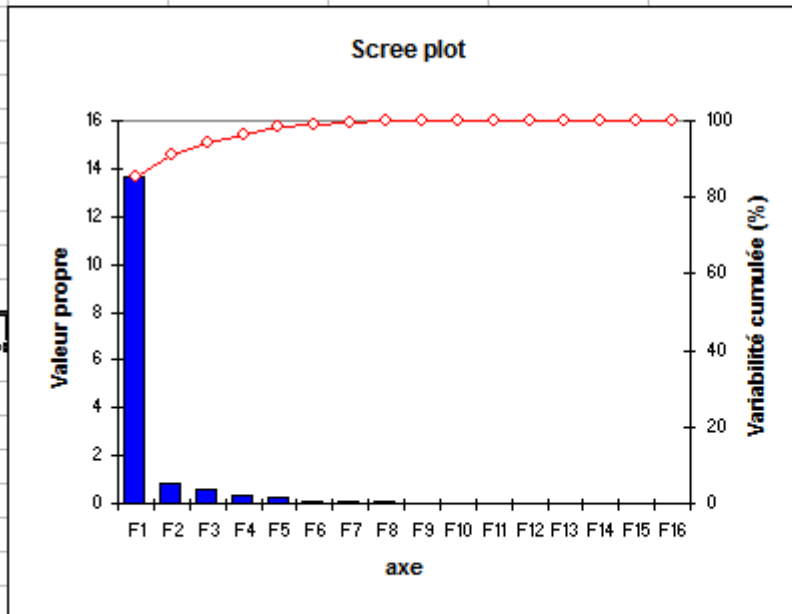
- Menu « Analyse des Données »
- ACP



Analyse en Composantes Principales :

Valeurs propres :

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	13.683	0.856	0.549	0.356	0.270	0.112	0.0
Variabilité (%)	85.521	5.348	3.429	2.227	1.685	0.699	0.5
% cumulé	85.521	90.869	94.298	96.525	98.210	98.909	99.4



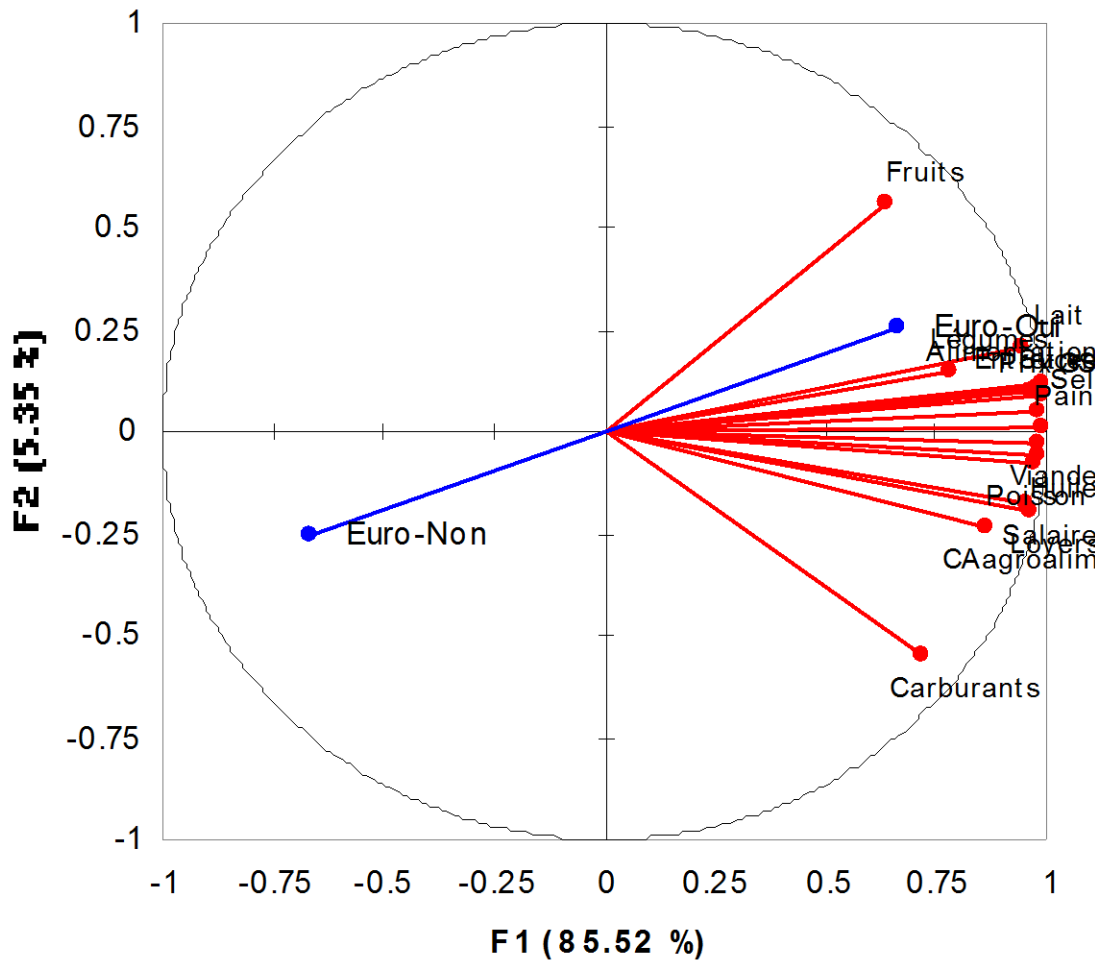
Vecteurs propres :

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Alimentation	0.268	0.130	0.022	0.060	-0.033	-0.030	-0.058	0.178	
Pain	0.268	0.006	-0.042	-0.141	0.008	-0.070	0.230	0.009	
Viande	0.266	-0.032	0.031	0.019	0.015	-0.253	-0.169	0.608	
Poisson	0.264	-0.082	-0.029	-0.017	-0.088	-0.114	-0.539	-0.372	
Lait	0.255	0.218	-0.139	-0.290	-0.250	0.245	-0.258	0.198	
Huile	0.265	-0.059	0.015	0.029	-0.011	-0.270	-0.211	-0.538	
Fruits	0.172	0.607	0.667	0.257	0.192	0.165	-0.013	-0.073	
Légumes	0.213	0.163	-0.494	0.761	-0.228	0.159	0.115	-0.023	
Sucre	0.260	0.102	-0.111	-0.264	0.017	0.117	0.606	-0.199	
Sel	0.266	0.048	-0.024	-0.233	-0.115	0.131	0.109	-0.133	
Salaire	0.258	-0.195	0.112	0.147	0.199	-0.453	0.281	0.000	
Loyers	0.261	-0.208	0.113	0.126	0.106	-0.285	0.057	0.138	
CAagroalim	0.233	-0.256	-0.175	0.029	0.766	0.482	-0.155	0.018	
Prix GS	0.267	0.090	-0.064	-0.165	-0.093	0.010	-0.091	0.210	
Entretien	0.266	0.113	-0.099	-0.187	-0.063	-0.001	0.043	-0.065	
Carburants	0.194	-0.591	0.452	0.145	-0.420	0.426	0.080	0.021	

Coordonnées des variables :

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Alimentation	0.991	0.120	0.016	0.036	-0.017	-0.010	-0.017	0.043	
Pain	0.993	0.006	-0.031	-0.084	0.004	-0.023	0.068	0.002	
Viande	0.983	-0.030	0.023	0.011	0.008	-0.085	-0.050	0.145	
Poisson	0.975	-0.076	-0.021	-0.010	-0.046	-0.038	-0.160	-0.089	
Lait	0.942	0.202	-0.103	-0.173	-0.130	0.082	-0.077	0.047	
Huile	0.980	-0.054	0.011	0.017	-0.006	-0.090	-0.063	-0.128	
Fruits	0.636	0.561	0.494	0.153	0.100	0.055	-0.004	-0.018	
Légumes	0.787	0.151	-0.366	0.454	-0.118	0.053	0.034	-0.005	
Sucre	0.960	0.094	-0.082	-0.158	0.009	0.039	0.180	-0.048	
Sel	0.984	0.044	-0.018	-0.139	-0.060	0.044	0.032	-0.032	
Salaire	0.955	-0.181	0.083	0.088	0.103	-0.151	0.083	0.000	
Loyers	0.966	-0.192	0.084	0.075	0.055	-0.095	0.017	0.033	
CAagroalim	0.861	-0.237	-0.130	0.017	0.398	0.161	-0.046	0.004	
Prix GS	0.987	0.084	-0.048	-0.099	-0.048	0.003	-0.027	0.050	
Entretien	0.984	0.105	-0.073	-0.112	-0.033	0.000	0.013	-0.015	
Carburants	0.716	-0.546	0.335	0.087	-0.218	0.143	0.024	0.005	
Euro-Non	-0.661	-0.250	0.058	0.058	-0.107	0.211	-0.249	0.011	
Euro-Oui	0.661	0.250	-0.058	-0.058	0.107	-0.211	0.249	-0.011	

Variables (axes F1 et F2 : 90.87 %)



● variables actives ● Variables supplémentaires

Contributions des variables (%) :

	F1	F2
Alimentation	7.173	1.677
Pain	7.201	0.004
Viande	7.061	0.102
Poisson	6.952	0.670
Lait	6.478	4.766
Huile	7.018	0.343
Fruits	2.952	36.818
Légumes	4.525	2.652
Sucre	6.737	1.043
Sel	7.069	0.228
Salaire	6.660	3.812
Loyers	6.815	4.328
CAagroalim	5.412	6.560
Prix GS	7.126	0.818
Entretien	7.073	1.277
Carburants	3.748	34.902

Observation	F1	F2
janv-98	-5.679	-0.109
févr-98	-5.975	1.161
02-juil	4.128	-0.723
03-juil	4.752	-0.793
Euro-Non	-3.459	-0.327
Euro-Oui	1.729	0.164

Observations (axes F1 et F2 : 90.87 %)

