

## DEUX EXEMPLES D'ACP

### EXEMPLE 1

On étudie les consommations annuelles en 1972, exprimées en francs, de 8 denrées alimentaires (les variables), les individus étant 8 catégories socio-professionnelles. Les données sont des moyennes par CSP :

	PAO	PAA	VIO	VIA	POT	LEC	RAI	PLP
AGRI	167	1	163	23	41	8	6	6
SAAG	162	2	141	12	40	12	4	15
PRIN	119	6	69	56	39	5	13	41
CSUP	87	11	63	111	27	3	18	39
CMOY	103	5	68	77	32	4	11	30
EMPL	111	4	72	66	34	6	10	28
OUVR	130	3	76	52	43	7	7	16
INAC	138	7	117	74	53	8	12	20

AGRI = Exploitants agricoles  
 SAAG= Salariés agricoles  
 PRIN = Professions indépendantes  
 CSUP = Cadres supérieurs  
 CMOY= Cadres moyens  
 EMPL= Employés  
 OUVR = Ouvriers  
 INAC = Inactifs

PAO = Pain ordinaire  
 PAA = Autre pain  
 VIO = Vin ordinaire  
 VIA=Autre vin  
 POT= Pommes de terre  
 LEC=Légumes secs  
 RAI=Raisin de tables  
 PLP= Plats préparés

Le programme SAS permettant d'obtenir les sorties ci-dessous est en Annexe.

#### The PRINCOMP Procedure

```
Observations      8
Variables          8
```

**Tableau 1**

#### Correlation Matrix

	PAO	PAA	VIO	VIA	POT	LEC	RAI	PLP
PAO	1.0000	-.7737	0.9262	-.9058	0.6564	0.8886	-.8334	-.8558
PAA	-.7737	1.0000	-.6040	0.9044	-.3329	-.6734	0.9588	0.7712
VIO	0.9262	-.6040	1.0000	-.7502	0.5171	0.7917	-.6690	-.8280
VIA	-.9058	0.9044	-.7502	1.0000	-.4186	-.8386	0.9239	0.7198
POT	0.6564	-.3329	0.5171	-.4186	1.0000	0.6029	-.4099	-.5540
LEC	0.8886	-.6734	0.7917	-.8386	0.6029	1.0000	-.8245	-.7509
RAI	-.8334	0.9588	-.6690	0.9239	-.4099	-.8245	1.0000	0.8344
PLP	-.8558	0.7712	-.8280	0.7198	-.5540	-.7509	0.8344	1.0000

**Tableau 2**

#### Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	6.20794684	5.32826545	0.7760	0.7760
2	0.87968139	0.46372027	0.1100	0.8860
3	0.41596112	0.10950645	0.0520	0.9379
4	0.30645467	0.13801317	0.0383	0.9763
5	0.16844150	0.15037379	0.0211	0.9973
6	0.01806771	0.01462094	0.0023	0.9996
7	0.00344677	0.00344677	0.0004	1.0000
8	0.00000000		0.0000	1.0000

**Tableau 3**

#### Eigenvectors

Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
-------	-------	-------	-------	-------	-------	-------	-------

PAO	-.391311	0.137823	0.161714	0.119350	0.294045	-.397748	0.106920	0.728963
PAA	0.348674	0.440585	0.319950	0.217909	-.265442	-.520704	-.423079	-.117773
VIO	-.349193	0.201682	0.680632	-.028883	0.245716	0.464752	-.253923	-.180130
VIA	0.373625	0.260309	0.073482	-.396545	-.345605	0.422866	-.033345	0.575000
POT	-.246371	0.743826	-.557660	-.073992	0.175725	0.107747	-.093428	-.135449
LEC	-.364822	0.128021	0.032401	0.518889	-.669192	0.184942	0.313107	0.012735
RAI	0.373052	0.325980	0.254250	0.063706	0.271532	-.016265	0.765903	-.158952
PLP	0.361676	-.050227	-.161692	0.708103	0.332914	0.360245	-.224966	0.218851

**Tableau 4**

Coordonnees et qualite de representation des individus

csp	Prin1	Prin2	qlt1	qlt2
AGRI	-3.37158	-0.24582	0.88444	0.00470
SAAG	-3.52171	-0.44740	0.89806	0.01449
PRIN	1.47203	0.05851	0.57460	0.00091
CSUP	4.35879	0.17611	0.94182	0.00154
CMOY	1.71808	-0.85665	0.75288	0.18717
EMPL	0.80653	-0.80853	0.42778	0.42990
OVR	-0.89910	-0.18304	0.36060	0.01495
INAC	-0.56304	2.30681	0.05552	0.93193

**Tableau 5**

Coordonnees des variables sur les axes

Pearson Correlation Coefficients, N = 8

	Prin1	Prin2
PAO	-0.97498	0.12927
PAA	0.86875	0.41323
VIO	-0.87004	0.18916
VIA	0.93092	0.24415
POT	-0.61385	0.69764
LEC	-0.90898	0.12007
RAI	0.92949	0.30574
PLP	0.90114	-0.04711

## INTERPRETATION

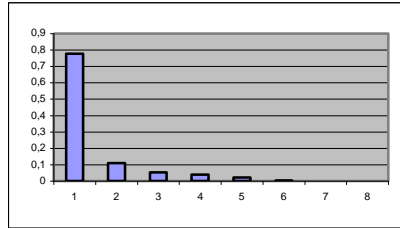
### 1) Choix du nombre d'axes à retenir

Nous utilisons pour cela le tableau 2. Le critère de Kaiser nous conduit à sélectionner un seul axe, qui retient 77% de l'inertie totale. L'axe 2 retient tout de même 11% de l'inertie, ce qui n'est pas négligeable, et qui conduit à un taux d'inertie expliquée de 89%, ce qui est un très bon résultat. Il peut être donc intéressant de l'étudier aussi. Nous le retiendrons si nous pouvons l'interpréter.

#### Remarque :

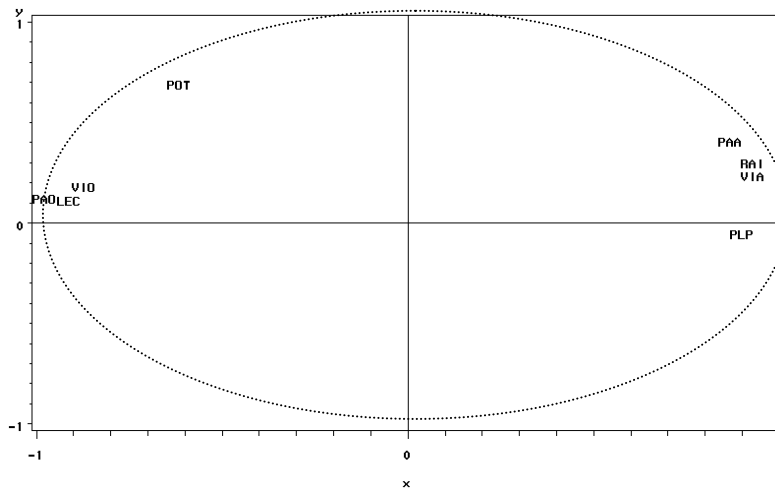
Le critère du coude ainsi que celui du Scree-test nous conduisent à retenir les deux premiers axes (voir ci-dessous).

Differences	Différences secondes
5,32826545	4,86454518
0,46372027	0,35421382
0,10950645	-0,02850672
0,13801317	-0,01236062
0,15037379	0,13575285
0,01462094	0,01117417

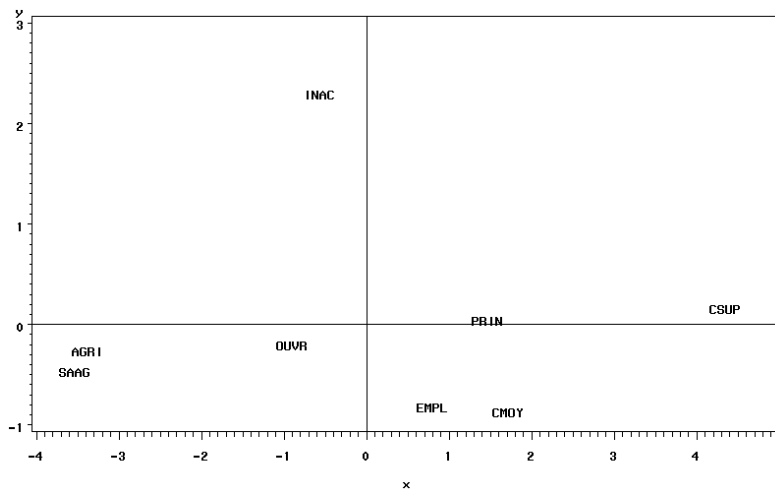


## 2) Graphes sur le plan factoriel (1,2)

Representation des variables axe2 \* axe1



Representation des individus axe2 \* axe1



## 3) Interprétation des axes

L'interprétation des axes factoriels se fait séquentiellement, pour chaque axe et chaque nuage de points, en regardant les contributions à la formation des axes.

### Axe 1 :

Variables : On sait que les variables contribuant le plus à la formation de l'axe 1 sont celles dont les coordonnées sur cet axe sont proches de 1 en valeur absolue. Pour repérer les contributions significatives, on

utilise le tableau 3 : on compare les valeurs de la colonne Prin1, coordonnées du premier axe factoriel, à la racine de la contribution moyenne  $1/\sqrt{8}=35\%$ , le signe donnant le sens de contribution. On obtient :

-	+
PAO LEC (VIO)	RAI VIA PLP (PAA)

PAA et VIO sont très proches de la contribution moyenne, on les intégrera donc dans l'interprétation de l'axe si elles vont dans le sens de l'interprétation que l'on peut en faire, sans elles.

L'axe 1 oppose les individus consommant du pain ordinaire, des légumes secs (et éventuellement du vin ordinaire) à ceux qui consomment du raisin, du vin (éventuellement du pain) plus sophistiqué et des plats préparés.

L'axe 1, et donc la première composante principale, mesure la répartition entre aliments ordinaires bon marchés et aliments plus recherchés.

Toutes les variables sont bien représentées sur l'axe (tableau 5 : la qualité de représentation est égale à la coordonnée au carré ; ainsi, la qualité de représentation de la variable PAO est égale à  $(-0,97^2)=0,94$  ; plus simplement, on peut avoir une idée de la qualité de représentation d'une variable sur un axe en lisant directement le graphique : une variable bien représentée est proche du bord du cercle des corrélations et à proximité de l'axe). La première composante principale explique donc correctement tous les types de consommations alimentaires.

Individus : De même, les individus contribuant le plus à la formation de l'axe 1 sont ceux dont les coordonnées sur cet axe sont les plus élevées en valeur absolue. Pour repérer les contributions significatives, on utilise le tableau 4 : on compare les valeurs de la colonne Prin1, coordonnées des individus, à la racine carrée de la première valeur propre  $\sqrt{6,2079}=2,49$ , le signe donnant le sens de contribution. On obtient :

-	+
AGRI SAAG	CSUP

Le premier axe met donc en opposition quant à leurs habitudes alimentaires les agriculteurs et les cadres supérieurs.

Les autres catégories socio-professionnelles, assez bien représentées sur l'axe à l'exception des inactifs (cf. contributions des individus sur l'axe 1), s'échelonnent suivant la hiérarchie habituelle. Elles sont bien expliquées par l'axe.

Conclusion : L'axe 1 reflète donc l'opposition qui existe entre les catégories socio-professionnelles dans leur alimentation, opposant les CSP modestes qui consomment des produits basiques aux catégories favorisées qui consomment des produits plus recherchés.

## Axe 2 :

Variables : Dans le tableau 3, on compare les valeurs de la colonne Prin2 à 35%, le signe donnant le sens de contribution. On obtient :

-	+
	POT PAA

L'axe 2 est défini par les variables POT et PAA. Compte tenu de la différence de contribution existant entre ces deux variables, de la contribution élevée de POT (55%), et de la qualité de représentation moyenne de PAA, la deuxième composante principale peut être considérée comme essentiellement liée à la consommation de pommes de terre.

Les variables, à l'exception de POT et de PAA (dans une moindre mesure) sont assez mal représentées sur l'axe (tableau 5). La deuxième composante principale n'explique donc qu'un aspect très particulier de la consommation alimentaire.

**Individus** : Pour repérer les individus ayant une contribution significative, on utilise le tableau 4 : on compare les valeurs de la colonne Prin2, coordonnées des individus sur l'axe 2, à la racine de la deuxième valeur propre  $\sqrt{0,8797} = 0,94$ , le signe donnant le sens de contribution. On obtient :

-	+
	INAC

Le deuxième axe est caractéristique des inactifs (expliquant  $\frac{c_{i2}^2}{n\lambda_2} = 75\%$  de l'inertie de l'axe).

Les autres catégories socio-professionnelles sont mal représentées sur l'axe.

**Conclusion** : L'axe 2 reflète donc la particularité des inactifs quant à leur alimentation, fortement composée de pommes de terre (un retour aux données d'origine vient confirmer cette conclusion).

### **Synthèse** :

Ici, il n'y a rien de plus à tirer de cette analyse au vu du graphique (pas de regroupement particuliers de points, autres qu'à proximité des axes, ce qui a déjà été analysé). On peut en revanche synthétiser les résultats de l'analyse dans un tableau C « réduit », tableau contenant l'essentiel (88%) de l'inertie (i.e. de l'information) totale du tableau d'origine.

CSP	C1	C2
AGRI	-3.37158	-0.24582
SAAG	-3.52171	-0.44740
PRIN	1.47203	0.05851
CSUP	4.35879	0.17611
CMOY	1.71808	-0.85665
EMPL	0.80653	-0.80853
OUVR	-0.89910	-0.18304
INAC	-0.56304	2.30681

C1=répartition entre aliments ordinaires et aliments plus recherchés

C2=répartition de la consommation de pommes de terre

## **EXEMPLE 2**

Le tableau suivant fournit la structure du bilan d'un groupe pétrolier de 1969 à 1984 :

Année	NET	INT	SUB	LMT	DCT	IMM	EXP	VRD
1969	17.93	3.96	0.88	7.38	19.86	25.45	5.34	19.21
1970	16.21	3.93	0.94	9.82	19.11	26.58	5.01	18.40
1971	19.01	3.56	1.91	9.43	17.87	25.94	5.40	16.88
1972	18.05	3.33	1.73	9.72	18.83	26.05	5.08	17.21
1973	16.56	3.10	2.14	9.39	20.36	23.95	6.19	18.31
1974	13.09	2.64	2.44	8.10	25.05	19.48	11.61	17.59
1975	13.43	2.42	2.45	10.83	22.07	22.13	11.17	15.49
1976	9.83	2.46	1.79	11.81	24.10	22.39	11.31	16.30
1977	9.46	2.33	2.30	11.46	24.45	23.07	11.16	15.77
1978	10.93	2.95	2.25	10.72	23.16	24.17	9.64	16.20
1979	13.02	3.74	2.21	7.99	23.04	19.53	12.60	17.87
1980	13.43	3.60	2.29	7.09	23.59	17.61	16.67	15.72
1981	13.37	3.35	2.58	6.76	23.94	18.04	15.42	16.54
1982	11.75	2.74	3.11	7.37	25.04	18.11	14.71	17.18
1983	12.59	3.05	3.85	7.12	23.40	19.17	11.86	18.97
1984	13.00	3.00	4.00	7.00	24.00	20.00	12.00	17.00

Les postes de bilan sont les suivants :

NET : Situation nette ; représente l'ensemble des capitaux propres de l'entreprise.

INT : Intérêts ; représente l'ensemble des frais financiers supportés par l'entreprise.

SUB : Subventions ; représente le montant total des subventions accordées par l'Etat.  
LMT : Dettes à long et moyen terme.  
DCT : Dettes à court terme.  
IMM : Immobilisations ; représente l'ensemble des terrains et du matériel de l'entreprise.  
EXP : Valeurs d'exploitation.  
VRD : Valeurs réalisables et disponibles ; ensemble des créances à court terme de l'entreprise.

Les données ont été ventilées en pourcentage par année, la somme des éléments d'une même ligne vaut 100, de manière à éviter les effets dus à l'inflation. On cherche à répondre aux questions suivantes :

- Quelle a été l'évolution de la structure de bilan sur 15 ans ?
- Peut-on mettre en évidence plusieurs périodes ? Si oui, comment se caractérisent-elles ?

#### The PRINCOMP Procedure

Observations 16  
Variables 8

**Tableau 0 :**

Simple Statistics				
	NET	INT	SUB	LMT
Mean	13.85375000	3.135000000	2.304375000	8.874375000
Std	2.81018432	0.514575067	0.818084415	1.674100388

Simple Statistics				
	DCT	IMM	EXP	VRD
Mean	22.36687500	21.97937500	10.32312500	17.16500000
Std	2.29420389	3.05337409	3.72700402	1.11731822

**Tableau1 :**

	<u>Correlation Matrix</u>							
	NET	INT	SUB	LMT	DCT	IMM	EXP	VRD
NET	1.0000	0.6861	-.4614	-.2041	-.8908	0.5536	-.7045	0.4784
INT	0.6861	1.0000	-.4494	-.4600	-.6007	0.2455	-.3398	0.5296
SUB	-.4614	-.4494	1.0000	-.4093	0.6127	-.6932	0.6079	-.1421
LMT	-.2041	-.4600	-.4093	1.0000	-.1884	0.5976	-.3904	-.4195
DCT	-.8908	-.6007	0.6127	-.1884	1.0000	-.8168	0.8644	-.3535
IMM	0.5536	0.2455	-.6932	0.5976	-.8168	1.0000	-.9446	0.2021
EXP	-.7045	-.3398	0.6079	-.3904	0.8644	-.9446	1.0000	-.4614
VRD	0.4784	0.5296	-.1421	-.4195	-.3535	0.2021	-.4614	1.0000

**Tableau2 :**

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	4.47037150	2.35552573	0.5588	0.5588
2	2.11484576	1.43418677	0.2644	0.8232
3	0.68065899	0.17991239	0.0851	0.9082
4	0.50074660	0.34116829	0.0626	0.9708
5	0.15957831	0.09542998	0.0199	0.9908
6	0.06414833	0.05449844	0.0080	0.9988
7	0.00964990	0.00964928	0.0012	1.0000
8	0.00000062		0.0000	1.0000

**Tableau3 :**

	<u>Eigenvectors</u>							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8
NET	0.402088	0.238457	-.076441	-.472312	-.500342	0.075090	0.322932	0.438493
INT	0.297792	0.427526	-.452436	0.165899	0.628110	0.023067	0.307945	0.080227
SUB	-.351008	0.189653	0.498151	-.579440	0.473763	0.051376	0.116887	0.127148
LMT	0.094674	-.661253	0.059523	0.111232	0.189935	0.552888	0.356877	0.262189
DCT	-.451140	0.001154	0.063719	0.367209	-.129641	-.473781	0.537313	0.357195
IMM	0.410473	-.307837	0.150543	-.017412	0.270322	-.526713	-.374049	0.472619
EXP	-.437830	0.144302	-.362700	0.045132	-.019013	0.312604	-.469622	0.579022
VRD	0.231873	0.414183	0.616025	0.513900	-.066083	0.293067	-.103434	0.173649

**Tableau4 :**

Coordonnees et qualite de representation des individus sur les axes

annee	Prin1	Prin2	q1t1	q1t2
1969	3.55662	1.50535	0.78441	0.14052
1970	3.57546	-0.04273	0.93110	0.00013
1971	3.12027	-0.21808	0.83031	0.00406
1972	2.87553	-0.54758	0.89332	0.03239
1973	1.84936	0.02352	0.75517	0.00012
1974	-1.42432	0.32194	0.57269	0.02926
1975	-0.79476	-1.97215	0.11144	0.68621
1976	-1.16070	-2.50400	0.15851	0.73770
1977	-1.59726	-2.65758	0.25931	0.71786
1978	-0.37918	-1.74803	0.03739	0.79463
1979	-0.36150	1.35612	0.04004	0.56350
1980	-1.75965	1.20307	0.34868	0.16299
1981	-1.75001	1.40025	0.49152	0.31468
1982	-2.51840	0.84115	0.87166	0.09724
1983	-1.37918	1.88579	0.21797	0.40752
1984	-1.85228	1.15298	0.50000	0.19373

**Tableau5 :**

Coordonnees des variables sur les axes

Pearson Correlation Coefficients, N = 16

	Prin1	Prin2
NET	0.85014	0.34678
INT	0.62963	0.62173
SUB	-0.74214	0.27580
LMT	0.20017	-0.96163
DCT	-0.95386	0.00168
IMM	0.86787	-0.44767
EXP	-0.92571	0.20985
VRD	0.49025	0.60233

## INTERPRETATION

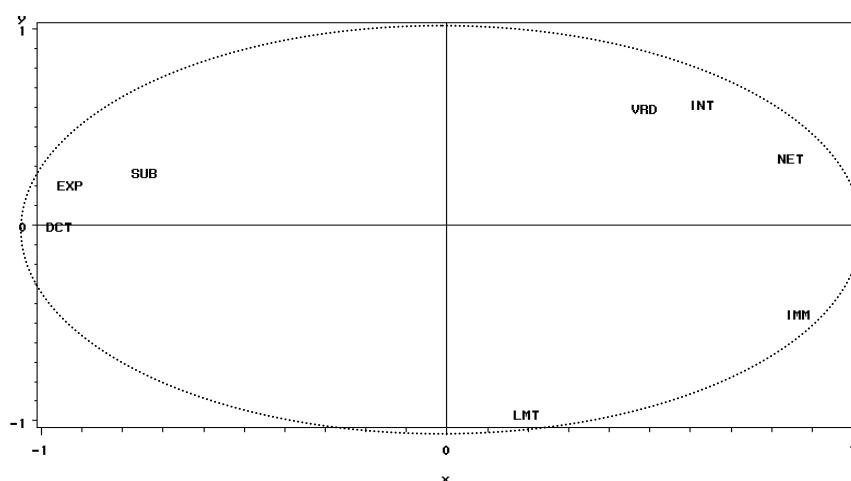
### 1) Choix du nombre d'axes à retenir

Nous utilisons pour cela le tableau 2. Le critère de Kaiser nous conduit à sélectionner deux axes, expliquant 82% de l'inertie totale du nuage.

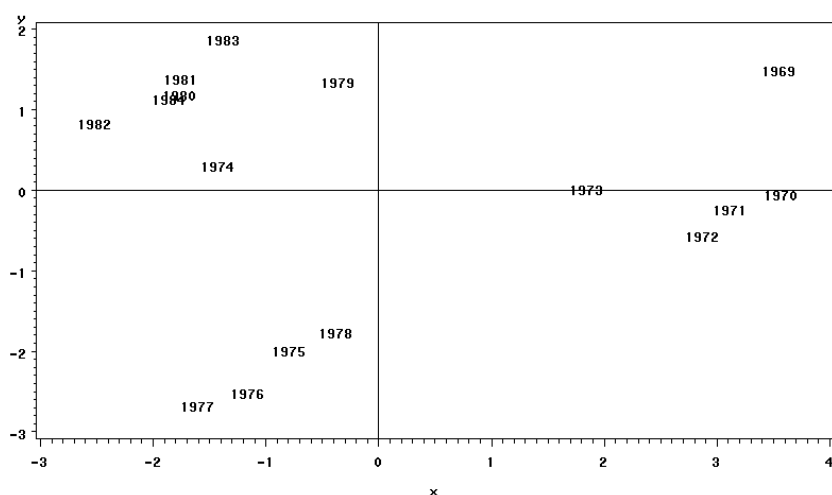
Le critère du coude conduit ici à sélectionner 4 axes ainsi que le Scree-test. En réalité, on peut ici se contenter de sélectionner deux axes. En effet, d'un coté, la chute d'inertie est très importante dès le troisième axe, qui ne conserve plus que 8% de l'inertie totale ; de l'autre, les deux premiers axes conservent plus de 80% de l'inertie ce qui est très bon. Avec les deux premiers axes, nous disposons donc d'un espace compréhensible par l'œil sans subir une déformation trop prononcée du nuage.

### 2) Graphes sur le plan factoriel (1,2)

Représentation des variables axe2 \* axe1



Représentation des individus axe2 \* axe1



### 3) Interprétation des axes

#### Axe 1 :

Variables : On utilise le tableau 3 : on compare les valeurs de la colonne Prin1, coordonnées du premier axe factoriel, à la racine de la contribution moyenne  $1/\sqrt{8}=35\%$ , le signe donnant le sens de contribution. On obtient :

-	+
DCT	NET
EXP	IMM
(SUB)	

SUB est très proche de la contribution moyenne, on l'intégrera dans l'interprétation si elle va dans le sens de celle qui est faite sans elle.

L'axe 1 oppose les postes DCT, EXP (dettes à court terme et valeur d'exploitation élevée) et éventuellement SUB (subventions) aux postes NET et IMM (capitaux propres élevés et nombre important d'immobilisation).

Toutes les variables sont bien représentées sur l'axe excepté LMT ( et VRD dans une moindre mesure) (tableau 5). La première composante principale explique donc correctement tous les postes, sauf ces deux derniers.



Individus : On utilise le tableau 4 : on compare les valeurs de la colonne Prin1 à la racine de la première valeur propre  $\sqrt{4,470} = 2,11$ , le signe donnant le sens de contribution. On obtient :

-	+
1982	1969 1970 1971 1972

Le premier axe met donc en opposition la structure de bilan l'année 1982 aux quatre années précédant le premier choc pétrolier. En fait, il isole les secondes, les autres coordonnées étant toutes négatives.

Sont bien représentées sur l'axe les années 73 74 **81 84**, en plus des années contribuant fortement à l'axe.

Conclusion : L'axe 1 oppose donc les années 1969 à 1972, marquées par un poids important dans la structure de leur bilan des postes NET et IMM, et un poids faible des postes DCT et EXP aux années plus récentes (à l'exception de 73), qui présentent le profil inverse. Pour illustrer ce résultat, nous pouvons revenir aux données sources. En nous servant des indications du tableau 0, nous avons :

Année	NET	DCT	IMM	EXP
1969	17.93	19.86	25.45	5.34
1970	16.21	19.11	<b>26.58</b>	<b>5.01</b>
1971	<b>19.01</b>	<b>17.87</b>	25.94	5.40
1972	18.05	18.83	26.05	5.08
1973	16.56	20.36	23.95	6.19
1982	11.75	25.04	18.11	14.71
Moyenne	13.85	22.37	21.98	10.32
Minimum	9.46	<b>17.87</b>	17.61	<b>5.01</b>
maximum	<b>19.01</b>	25.05	<b>26.58</b>	16.67

## Axe 2 :

Variables : Dans le tableau 3, on compare les valeurs de la colonne Prin2 à 35% le signe donnant le sens de contribution. On obtient :

-	+
LMT	INT VRD

L'axe 2 oppose le poste LMT (endettement à moyen et long terme) aux postes INT et VRD (créances à court terme et capitaux propres).

Les autres variables sont assez mal représentées sur l'axe, donc assez mal expliquées par celui-ci.

Individus : On utilise le tableau 4 : on compare les valeurs de la colonne Prin2, à  $\sqrt{2,11} = 1,45$ , le signe donnant le sens de contribution. On obtient :

-	+
1975 1976 1977 1978	1983 (1969)

L'axe 2 oppose les années 75 à 78 aux années 83 et éventuellement 69. Les années 75 à 78 sont celles de l'entre deux chocs pétroliers. Les années 83, et plus encore 69 sont assez mal représentées sur l'axe. Elles ne sont pas bien expliquées par l'axe ; on ne les fera donc pas intervenir dans l'interprétation.

Conclusion : L'axe 2 isole les années de l'entre deux chocs, 75 à 78, caractérisées par un poids important du poste LMT et un poids faible des postes VRD et INT. La encore, un retour aux données brutes vient confirmer cette interprétation :

Année	INT	LMT	VRD
1975	2.42	10.83	<b>15.49</b>
1976	2.46	<b>11.81</b>	16.30
1977	<b>2.33</b>	11.46	15.77
1978	2.95	10.72	16.20
1983	3.05	7.12	18.97
moyenne	3.13	8.87	17.16
minimum	<b>2.33</b>	6.76	<b>15.49</b>
maximum	3.96	<b>11.81</b>	19.21

### Synthèse :

L'interprétation de l'ACP peut être achevée par une analyse des proximités entre points sur les plans factoriels. En effet, au vu des graphiques, on peut distinguer plusieurs groupes. Avant d'interpréter ces proximités, il s'agit de s'assurer de la qualité de représentation des points :

Les qualités de représentation sur le plan (1,2) des individus sont données ci-dessous :

1969	0,92493
1970	0,93123
1971	0,83437
1972	0,92571
1973	0,75529
1974	0,60195
1975	0,79765
1976	0,89621
1977	0,97717
1978	0,83202
1979	0,60354
1980	0,51167
1981	0,8062
1982	0,9689
1983	0,62549
1984	0,69373

Les individus sont donc tous à peu près bien représentés sur les axes.

Il n'est pas nécessaire de calculer les qualités de représentation des variables. Les variables bien représentées sont celles qui sont proches du bord du cercle des corrélations (telles que  $x^2 + y^2 = 1$  ; en particulier les variables qui sont situées sur les axes et de coordonnée 1).

Ici, schématiquement, trois périodes apparaissent :

- la période antérieure au premier choc pétrolier 1969-1973. Le bilan est alors marqué par les postes situation nette et immobilisations. En revanche, l'endettement à court terme et le poste valeurs d'exploitation sont peu importants.
- Le période de l'entre deux chocs, 1975-1978, au cours de laquelle le groupe s'engage dans une politique d'endettement à long et moyen terme.
- La période postérieure au second choc, après 1979, qui voit l'essor de l'endettement à court terme, ainsi que dans une moindre mesure des subventions.

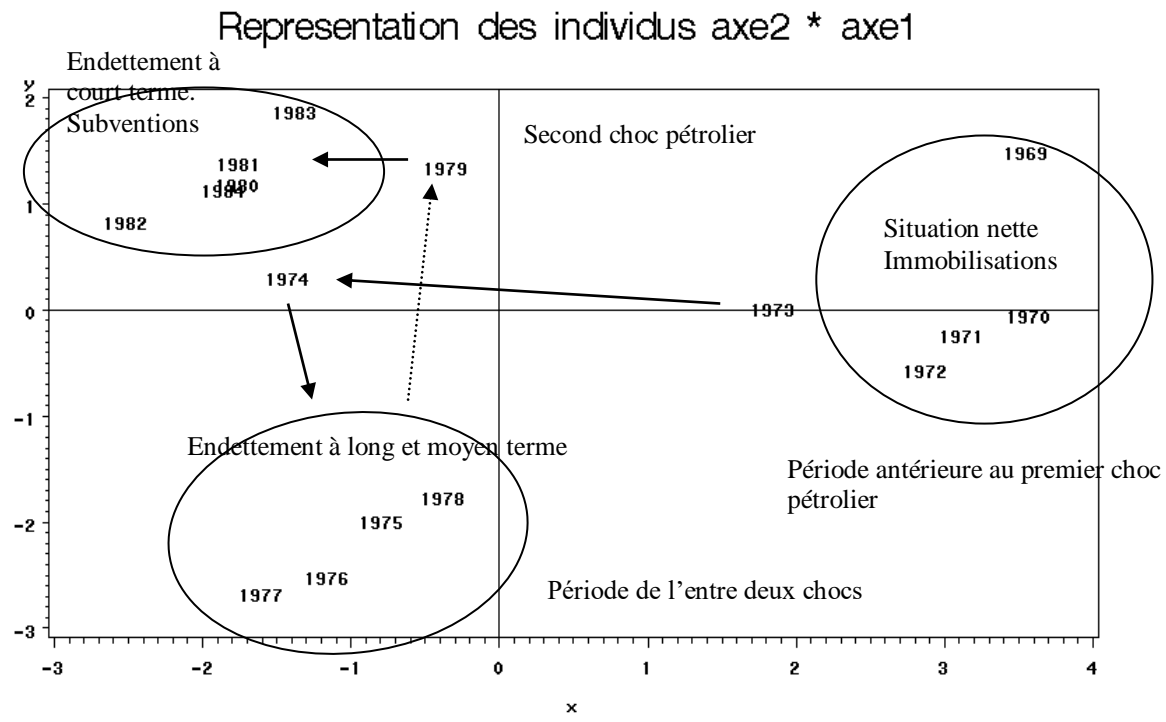
On peut également donner le tableau C , constitué des deux premières colonnes du tableau 4.

Conclusion : Sur cet exemple aux dimensions modestes, apparaissent les résultats que l'analyse des données permet d'obtenir :

- une mise en évidence des traits majeurs contenus dans les données et qui sont sériés axe par axe.
- l'existence de relations d'attraction ou d'opposition entre variables.
- une mise en évidence de sous-groupes homogènes que l'analyse permet d'interpréter.

Ici, la partition peut être opérée à vue sur le plan correspondant dans la mesure où le nombre d'éléments à représenter étant faible, le nuage ne présente pas de continuum de points. Tel n'est pas le cas lorsque le jeu de données est plus important.

Signalons que le plan précédent livré avec les commentaires appropriés constitue un excellent outil d'analyse. Cet exemple permet de souligner les écueils à éviter. Il importe de rendre les représentations graphiques lisibles et de les accompagner d'un texte expliquant à la fois leur obtention et leur signification.



## ANNEXE

```
data alim; input csp $ PAO PAA VIO VIA POT LEC RAI PLP;
cards;
AGRI 167 1 163 23 41 8 6 6
SAAG 162 2 141 12 40 12 4 15
PRIN 119 6 69 56 39 5 13 41
CSUP 87 11 63 111 27 3 18 39
CMOY 103 5 68 77 32 4 11 30
EMPL 111 4 72 66 34 6 10 28
OUVR 130 3 76 52 43 7 7 16
INAC 138 7 117 74 53 8 12 20;
run;
```

```
proc princomp data=alim out=b vardef=n;
var PAO PAA VIO VIA POT LEC RAI PLP;
run;
```

```
data c; set b;
array k{*} prin1-prin8;
disto=uss(of k{*});
qlt1=prin1*prin1/disto;
qlt2=prin2*prin2/disto;
```

```

keep prin1-prin2 q1t1-q1t2 csp;
run ;
proc print data=c;
id csp;
var prin1-prin2 q1t1-q1t2;
title 'Coordonnees et qualite de representation des individus sur les axes' ;
run ;

proc corr data=b out=d noprob nosimple;
var prin1-prin2;
with PAO PAA VIO VIA POT LEC RAI PLP;
run ;

data indiv;set b;
x=prin1;y=prin2;
xsys='2';ysys='2';
text = csp;
run;
proc gplot data=indiv;
title " Representation des individus axe2 * axe1";
plot y*x /annotate=indiv frame href=0 vref=0 ;
symbol1 v=none;
run;

data varia;set d;
x=prin1;y=prin2;xsys='2';ysys='2';
xmin=-1;xmax=1;ymax=1;ymin=-1;
text = _name_;
run;

proc gplot data=varia;
title " Representation des variables axe2 * axe1";
plot y*x /annotate=varia frame href=0 vref=0 haxis=-1 to 1 vaxis=-1 to 1 ;
symbol1 v=none;
run;
quit;

```

