3 Analyse Factorielle des Correspondances

3.1 Introduction

L'analyse factorielle des correspondances (AFC), ou analyse des correspondances simples, est une méthode exploratoire d'analyse des tableaux de contingence. Elle a été développé essentiellement par J.-P. Benzecri durant la période 1970-1990.

Soient deux variables nominales X et Y, comportant respectivement p et q modalités. On a observé les valeurs de ces variables sur une population et on dispose d'un tableau de contingence à p lignes et q colonnes donnant les effectifs conjoints c'est-à-dire les effectifs observés pour chaque combinaison d'une modalité i de X et d'une modalité j de Y.
Les valeurs de ce tableau seront notées n_{ij}, l'effectif total sera noté N.

L'ACP vise à analyser ce tableau en apportant des réponses à des questions telles que :

- Y a-t-il des lignes du tableau (modalités de X) qui se "ressemblent", c'est-à-dire telles que les distributions des modalités de Y soient analogues ?
- Y a-t-il des lignes du tableau (modalités de X) qui s'opposent, c'est-à-dire telles que les distributions des modalités de Y soient très différentes ?
- Mêmes questions pour les colonnes du tableau.
- Y a-t-il des associations modalité de X - modalité de Y qui s'attirent (effectif conjoint particulièrement élevé) ou qui se repoussent (effectif conjoint particulièrement faible) ?

La méthode se fixe également comme but de construire des représentations graphiques mettant en évidence ces propriétés des données.

3.2 Exemple

3.2.1 Enoncé

Réf. Examen de Statistiques de mai 2004, Module MULT, Maîtrise de Psychologie, Université René Descartes. Site Web : http://piaget.psycho.univ-paris5.fr/Statistiques/

Les données qui suivent sont constituées par les résultats du premier tour des élections régionales de 2004 pour la région Ile de France. Pour chacun des huit départements de l'Ile de France (en lignes), on a les effectifs de suffrages pour chacune des huit listes candidates ainsi que les effectifs d'abstentions (en colonnes). L'objectif est d'analyser la structure des votes ainsi que les liaisons entre listes et départements. Voici les codes de désignation des départements et des listes :

Départements	Code
Paris (75)	PARI
Seine et Marne (77)	SMAR
Yvelines (78)	YVEL
Essonne (91)	ESSO
Hauts de Seine (92)	HTSS
Seine Saint-Denis (93)	STDE
Val de Marne (94)	VDMA
Val d'Oise (95)	VDOI

Listes	Tête de liste	Code
PS-Verts-MRG-MRC	Huchon	HUCH
UMP	Copé	COPE
UDF	Santini	SANT
FN	Le Pen	LEPE
PC-AGR-AC	Buffet	BUFF
LO-LCR	Laguiller	LAGU
GE-Les Bleus	Pelegrin	PELE
MNR	Bay	BAY
Abstentions		ABST

Données : résultats du premier tour des régionales 2004 en Ile de France

	HUCHON	COPE	SANTINI	LEPEN	BUFFET	LAGU	PELEG	BAY	ABSTEN	TOTAL
PARI	258495	184419	114222	57183	39052	22479	13277	5006	434078	1128211
SMAR	128715	114003	48782	71897	25732	19738	11980	7085	301478	729410
YVEL	150141	140634	96746	61676	23292	15998	13939	6486	329626	838538
ESSO	144581	95451	59967	54309	26732	17545	12108	5346	270414	686453
HTSS	143444	136677	122610	47279	32987	16438	11322	4690	314964	830411
STDE	107327	61507	40081	54412	49535	19619	8393	5176	287618	633668
VDMA	126569	93049	60234	47074	41897	17308	10969	4557	286913	688570
VDOI	111176	82524	47903	55165	24693	17018	9876	4825	262458	615638
TOTAL	1170448	908264	590545	448995	263920	146143	91864	43171	2487549	6150899

Y a-t-il des départements qui se ressemblent, c'est-à-dire dans lesquels les résultats (en pourcentages) des différentes listes sont voisins ? Y a-t-il au contraire des départements qui s'opposent (résultats très différents) ?

Y a-t-il des départements dont les résultats sont proches de ceux de la région tout entière ? Y a-t-il des départements "à part" (dont les résultats s'écartent notablement de ceux de la région) ?

Y a-t-il des listes qui se ressemblent : elles n'obtiennent pas nécessairement les mêmes scores, mais les départements où elles obtiennent de bons scores sont les mêmes ? Y a-t-il des listes qui s'opposent ?

Y a-t-il des listes dont l'audience est la même dans tous les départements ? Y a-t-il des listes pour lesquelles le vote est concentré dans certains départements ?

Comment les départements "à part" et les listes à "vote concentré" s'associent-ils ?

3.2.2 Etude descriptive du tableau de contingence

On fixe les notations suivantes :

n_{ij} : effectif de la cellule (i,j),

$n_{i\bullet}$: effectif total de la ligne i,

$n_{\bullet j}$: effectif total de la colonne j

$n_{\bullet\bullet}$: effectif total

3.2.2.1 Tableau des fréquences

Les fréquences sont calculées par : $f_{ij} = \dfrac{n_{ij}}{n_{\bullet\bullet}} = \dfrac{Effectif\ \ de\ \ la\ \ cellule\ \ (i,j)}{Effectif\ \ \ total}$

	HUCHON	COPE	SANTINI	LEPEN	BUFFET	LAGU	PELEG	BAY	ABSTEN	TOTAL
PARI	4,20%	3,00%	1,86%	0,93%	0,63%	0,37%	0,22%	0,08%	7,06%	18,34%
SMAR	2,09%	1,85%	0,79%	1,17%	0,42%	0,32%	0,19%	0,12%	4,90%	11,86%
YVEL	2,44%	2,29%	1,57%	1,00%	0,38%	0,26%	0,23%	0,11%	5,36%	13,63%
ESSO	2,35%	1,55%	0,97%	0,88%	0,43%	0,29%	0,20%	0,09%	4,40%	11,16%
HTSS	2,33%	2,22%	1,99%	0,77%	0,54%	0,27%	0,18%	0,08%	5,12%	13,50%
STDE	1,74%	1,00%	0,65%	0,88%	0,81%	0,32%	0,14%	0,08%	4,68%	10,30%
VDMA	2,06%	1,51%	0,10%	0,77%	0,68%	0,28%	0,18%	0,07%	4,66%	11,19%
VDOI	1,81%	1,34%	0,78%	0,90%	0,40%	0,28%	0,16%	0,08%	4,27%	10,01%
TOTAL	19,03%	14,77%	9,60%	7,30%	4,29%	2,38%	1,49%	0,70%	40,44%	100,00%

3.2.2.2 Tableau des fréquences lignes

Les fréquences lignes (ou coordonnées des profils lignes) sont calculées par :

$$fl_{ij} = \frac{n_{ij}}{n_{i\bullet}} = \frac{f_{ij}}{f_{i\bullet}} = \frac{Effectif\ \ de\ \ la\ \ cellule\ \ (i,j)}{Effectif\ \ de\ \ la\ \ ligne\ \ i}$$

Les coordonnées du profil ligne moyen sont calculées par : $f_{\bullet j} = \dfrac{n_{\bullet j}}{n_{\bullet\bullet}} = \dfrac{Effectif\ \ de\ la\ \ colonne\ \ j}{Effectif\ \ \ total}$

	HUCHON	COPE	SANTINI	LEPEN	BUFFET	LAGU	PELEG	BAY	ABSTEN	TOTAL
PARI	22,91%	16,35%	10,12%	5,07%	3,46%	1,99%	1,18%	0,44%	38,47%	100,00%
SMAR	17,65%	15,63%	6,69%	9,86%	3,53%	2,71%	1,64%	0,97%	41,33%	100,00%
YVEL	17,91%	16,77%	11,54%	7,36%	2,78%	1,91%	1,66%	0,77%	39,31%	100,00%
ESSO	21,06%	13,90%	8,74%	7,91%	3,89%	2,56%	1,76%	0,78%	39,39%	100,00%
HTSS	17,27%	16,46%	14,76%	5,69%	3,97%	1,98%	1,36%	0,56%	37,93%	100,00%
STDE	16,94%	9,71%	6,33%	8,59%	7,82%	3,10%	1,32%	0,82%	45,39%	100,00%
VDMA	18,38%	13,51%	8,75%	6,84%	6,08%	2,51%	1,59%	0,66%	41,67%	100,00%
VDOI	18,06%	13,40%	7,78%	8,96%	4,01%	2,76%	1,60%	0,78%	42,63%	100,00%
TOTAL	19,03%	14,77%	9,60%	7,30%	4,29%	2,38%	1,49%	0,70%	40,44%	100,00%

3.2.2.3 Tableau des fréquences colonnes

Les fréquences colonnes (ou coordonnées des profils colonnes) sont calculées par :

$$fc_{ij} = \frac{n_{ij}}{n_{\bullet j}} = \frac{f_{ij}}{f_{\bullet j}} = \frac{Effectif\ \ de\ \ la\ \ cellule\ \ (i,j)}{Effectif\ \ de\ \ la\ \ colonne\ \ j}$$

Les coordonnées du profil colonne moyen sont calculées par : $f_{i\bullet} = \dfrac{n_{i\bullet}}{n_{\bullet\bullet}} = \dfrac{Effectif\ \ de\ \ la\ \ ligne\ \ i}{Effectif\ \ \ total}$

	HUCHON	COPE	SANTINI	LEPEN	BUFFET	LAGU	PELEG	BAY	ABSTEN	TOTAL
PARI	22,09%	20,30%	19,34%	12,74%	14,80%	15,38%	14,45%	11,60%	17,45%	18,34%
SMAR	11,00%	12,55%	8,26%	16,01%	9,75%	13,51%	13,04%	16,41%	12,12%	11,86%
YVEL	12,83%	15,48%	16,38%	13,74%	8,83%	10,95%	15,17%	15,02%	13,25%	13,63%
ESSO	12,35%	10,51%	10,15%	12,10%	10,13%	12,01%	13,18%	12,38%	10,87%	11,16%
HTSS	12,26%	15,05%	20,76%	10,53%	12,50%	11,25%	12,32%	10,86%	12,66%	13,50%
STDE	9,17%	6,77%	6,79%	12,12%	18,77%	13,42%	9,14%	11,99%	11,56%	10,30%
VDMA	10,81%	10,24%	10,20%	10,48%	15,87%	11,84%	11,94%	10,56%	11,53%	11,19%
VDOI	9,50%	9,09%	8,11%	12,29%	9,36%	11,64%	10,75%	11,18%	10,55%	10,01%
TOTAL	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

3.2.2.4 Distances entre profils. Métrique du Φ^2

Chaque ligne du tableau des fréquences lignes peut être vue comme la liste des coordonnées d'un point dans un espace à q dimensions. On obtient ainsi le nuage des individus-lignes. On définit de même le nuage des individus-colonnes à partir du tableau des fréquences colonnes.

Comme en ACP, on s'intéresse alors aux directions de "plus grande dispersion" de chacun de ces nuages de points. Mais, pour mesurer la "distance" entre deux individus, on utilise la *métrique du* Φ^2 au lieu de la distance habituelle (dite *métrique euclidienne*). La distance du Φ^2 entre la ligne i et la ligne i' est ainsi définie par :

$$d^2_{\Phi^2}(L_i, L_{i'}) = \sum_j \frac{(fl_{ij} - fl_{i'j})^2}{f_{\bullet j}}$$

Pourquoi utiliser cette métrique plutôt que la métrique euclidienne ? Deux raisons fortes peuvent être avancées :

- Avec la métrique du Φ^2, la distance entre deux lignes ne dépend pas des poids respectifs des colonnes. Ainsi, sur notre exemple, les différentes listes obtiennent des scores très différents et l'usage de la métrique euclidienne aurait donné trop de poids aux listes qui ont obtenu des scores élevés (ABST, HUCH, COPE).

- La métrique du Φ^2 possède la propriété d'*équivalence distributionnelle* : si on regroupe deux modalités lignes, les distances entre les profils-colonne, ou entre les autres profils-lignes restent inchangées.

Par exemple, la distance entre la ligne PARI et la ligne SMAR est donnée par :

$$d^2_{\Phi^2}(PARI, SMAR) = \frac{(0{,}2291 - 0{,}1765)^2}{0{,}1903} + \ldots + \frac{(0{,}3847 - 0{,}4133)^2}{0{,}4044} = 0{,}0682$$

La distance entre PARI et le profil-ligne moyen est donnée par :

$$d^2_{\Phi^2}(PARI, Moyenne) = \frac{(0{,}2291 - 0{,}1903)^2}{0{,}1903} + \ldots + \frac{(0{,}3847 - 0{,}4044)^2}{0{,}4044} = 0{,}0215$$

Avec les transpositions nécessaires, ce qui vient d'être dit pour les lignes s'applique également aux colonnes. Par exemple, la distance entre la colonne BUFFET et la colonne SANTINI est :

$$d^2_{\Phi^2}(BUFFET, SANTINI) = \frac{(0{,}1480 - 0{,}1934)^2}{0{,}1834} + \ldots + \frac{(0{,}0936 - 0{,}0811)^2}{0{,}1001} = 0{,}2753$$

Notons qu'en revanche, il n'existe pas d'outil mesurant une "distance" entre une ligne et une colonne.

3.2.2.5 Taux de liaison et Phi-2

Les taux de liaison sont définis par : $t_{ij} = \dfrac{f_{ij} - f_{i\bullet} f_{\bullet j}}{f_{i\bullet} f_{\bullet j}}$

	HUCHON	COPE	SANTINI	LEPEN	BUFFET	LAGU	PELEG	BAY	ABSTEN
PARI	0,204	0,107	0,054	-0,306	-0,193	-0,161	-0,212	-0,368	-0,049
SMAR	-0,073	0,058	-0,303	0,350	-0,178	0,139	0,100	0,384	0,022
YVEL	-0,059	0,136	0,202	0,008	-0,353	-0,197	0,113	0,102	-0,028
ESSO	0,107	-0,058	-0,090	0,084	-0,092	0,076	0,181	0,110	-0,026
HTSS	-0,092	0,115	0,538	-0,220	-0,074	-0,167	-0,087	-0,195	-0,062
STDE	-0,110	-0,343	-0,341	0,176	0,822	0,303	-0,113	0,164	0,122
VDMA	-0,034	-0,085	-0,089	-0,063	0,418	0,058	0,067	-0,057	0,030
VDOI	-0,051	-0,092	-0,190	0,228	-0,065	0,163	0,074	0,117	0,054

Signification pratique du taux de liaison : le score de la liste Huchon à Paris est 20% plus élevé que le score théorique que l'on observerait si les votes étaient indépendants des départements. Au contraire, celui de la liste Le Pen est 30% moins élevé que le score théorique.

Par construction, les valeurs prises par le taux de liaison sont :

- des nombres positifs quelconques (un score observé peut être 200% ou 300% supérieur au score théorique)
- des nombres négatifs compris entre -1 et 0 (le "déficit" le plus extrême d'un score observé est d'être 100% moins élevé que le score théorique).

Notez que le coefficient $f_{i\bullet} f_{\bullet j}$ représente le "poids théorique" de chaque cellule dans le tableau. La somme de ces coefficients vaut 1.

La moyenne de la série des taux de liaison pondérée par les coefficients $f_{i\bullet} f_{\bullet j}$ est nulle. La variance de cette série (avec les mêmes pondérations) est le coefficient Φ^2 :

$$\Phi^2 = \sum_{i,j} f_{i\bullet} f_{\bullet j} t_{ij}^2 = \sum_{i,j} \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})^2}{f_{i\bullet} f_{\bullet j}} = \frac{\mathrm{X}^2}{n_{\bullet\bullet}}$$

Ici, on obtient : $\Phi^2 = 0{,}02379$.

La méthode d'analyse factorielle des correspondances peut être vue comme une décomposition pertinente du Φ^2 selon plusieurs axes factoriels.

3.2.3 L'analyse factorielle des correspondances proprement dite

L'application de la méthode a deux effets :

- d'une part, on construit des images des nuages d'"individus-lignes" et d'"individus-colonnes" de départ, de façon que les distances entre images soient des distances euclidiennes et non plus des distances calculées selon la métrique du Φ^2;
- d'autre part, on recherche les directions de plus grande dispersion dans ces nuages de points images.

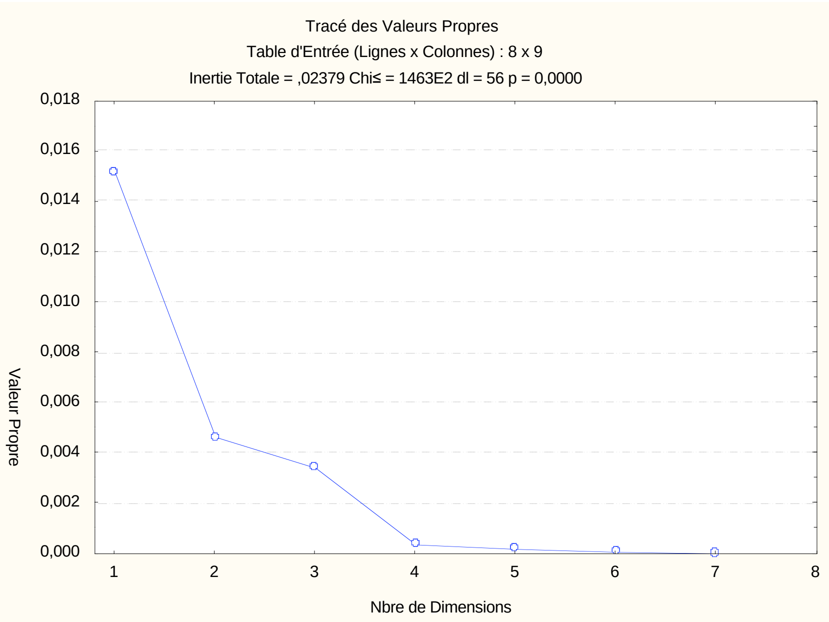
La matrice (tableau de valeurs) dont on recherche les valeurs propres et vecteurs propres est un objet mathématique "compliqué", qui ne possède pas de signification intuitive immédiate. De fait, on part de la matrice dont le terme à l'intersection de la ligne i et de la colonne j vaut : $\frac{f_{ij}}{\sqrt{f_{i\bullet}f_{\bullet j}}}$ et on calcule des produits scalaires entre lignes (ou entre colonnes) de cette matrice.

3.2.3.1 Valeurs propres

Le nombre de valeurs propres produites par la recherche des facteurs principaux est égal au minimum du nombre de lignes et du nombre de colonnes du tableau de contingence. Cependant, la première valeur propre est systématiquement égale à 1, et n'est pas mentionnée dans les résultats. Les autres valeurs propres sont des nombres positifs inférieurs à 1 et leur somme est égale à Φ^2.

Valeurs Propres et Inertie de toutes les Dimensions (idf.sta)
Inertie Totale = ,02379 Chi^2 = 146308 dl = 56 p = 0,0000

Dimension	ValSing.	ValProp.	%age Inertie	%age Cumulé	Chi^2
1	0,122976	0,015123	63,58	63,58	93020
2	0,068237	0,004656	19,58	83,15	28640
3	0,058363	0,003406	14,32	97,47	20951
4	0,018685	0,000349	1,47	98,94	2147
5	0,012321	0,000152	0,64	99,58	934
6	0,008701	0,000076	0,32	99,90	466
7	0,004936	0,000024	0,10	100,00	150



Le choix du nombre d'axes factoriels à conserver se fait comme dans le cas de l'ACP. Ici, on observe une brusque décroissance des valeurs propres entre la 3è et la 4è valeur propre. On retient donc les 3 premiers axes factoriels.

3.2.3.2 Résultats relatifs aux individus-lignes

Coordonnées Ligne et Contributions à l'Inertie (idf.sta)
Standardisation : Profils ligne et colonne

	Ligne N°	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos^2 Dim.1	Inertie Dim.2	Cos^2 Dim.2	Inertie Dim.3	Cos^2 Dim.3
PARI	1	-0,1050	0,0027	0,1016	0,1834	0,9924	0,1659	0,1337	0,5122	0,0003	0,0003	0,5561	0,4799
SMAR	2	0,0821	-0,1181	-0,0332	0,1186	0,9678	0,1122	0,0528	0,2993	0,3550	0,6194	0,0385	0,0491
YVEL	3	-0,0960	-0,0397	-0,0555	0,1363	0,9811	0,0810	0,0830	0,6517	0,0461	0,1113	0,1234	0,2181
ESSO	4	0,0183	-0,0393	0,0355	0,1116	0,5895	0,0249	0,0025	0,0627	0,0369	0,2898	0,0413	0,2369
HTSS	5	-0,1586	0,0824	-0,0752	0,1350	0,9966	0,2140	0,2245	0,6668	0,1967	0,1799	0,2239	0,1498
STDE	6	0,2478	0,0954	-0,0017	0,1030	0,9979	0,3061	0,4184	0,8691	0,2014	0,1288	0,0001	0,0000
VDMA	7	0,0706	0,0667	0,0115	0,1119	0,9225	0,0488	0,0369	0,4803	0,1071	0,4293	0,0044	0,0128
VDOI	8	0,0854	-0,0513	-0,0206	0,1001	0,9264	0,0470	0,0482	0,6529	0,0565	0,2356	0,0124	0,0379

Le tableau ci-dessus rassemble tous les résultats relatifs aux individus-lignes.

La colonne "Masse" rappelle les fréquences marginales des lignes c'est-à-dire le profil colonne moyen. Contrairement à l'ACP normée, dans laquelle chaque individu était affecté du même poids, les départements ont ici un "poids" dépendant de l'effectif total d'électeurs inscrits dans le département.

La colonne "Qualité" indique les qualités de représentation des individus ligne par les trois premiers axes principaux. Ces qualités sont calculées par des formules du type (Li désigne ici la ligne N°i, Fj, le facteur principal N°j) :

$$QLT(L_i, F_1; F_2; F_3) = \frac{(Coord\,de\,L_i\,selon\,F_1)^2 + (Coord\,de\,L_i\,selon\,F_2)^2 + (Coord\,de\,L_i\,selon\,F_3)^2}{\sum_l (Coord\,de\,L_i\,selon\,F_l)^2}$$

Par exemple :

$$QLT(PARIS, F_1; F_2; F_3) = \frac{(-0{,}1050)^2 + (0{,}0027)^2 + (0{,}1016)^2}{(-0{,}1050)^2 + (0{,}0027)^2 + (0{,}1016)^2 + (0{,}0107)^2 + (0{,}0068)^2 + (-0{,}0017)^2 + (-0{,}0007)^2}$$

La colonne "Inertie relative" est calculée de la manière suivante :

- L'inertie d'une combinaison individu-ligne individu-colonne correspondant à une cellule du tableau de contingence est le carré du taux de liaison, multiplié par la pondération (fréquence-ligne x fréquence colonne) correspondante.
- L'inertie absolue d'un individu-ligne est la somme des inerties des cellules de la ligne
- L'inertie relative d'un individu ligne est obtenue en divisant l'inertie absolue de l'individu par la somme de toutes les inerties, c'est-à-dire par Φ^2.

Pour chacun des trois axes factoriels, le tableau nous donne également les coordonnées ou *scores factoriels* de l'individu-ligne selon cet axe. Ces coordonnées ont les propriétés suivantes :

- Selon chaque axe, la moyenne des coordonnées des individus-lignes pondérées par les masses, est nulle.
- Selon chaque axe, la moyenne des carrés des coordonnées des individus-lignes pondérées par les masses, est égale à la valeur propre correspondante.
- Les coordonnées selon deux axes différents, pondérées par les masses, forment deux séries statistiques indépendantes (covariance nulle)

Ainsi :

$$(-0{,}1050\times 0{,}1834)+(0{,}0821\times 0{,}1186)+\ldots+(0{,}0854\times 0{,}1001)=0$$

$$(-0{,}1050)^2\times 0{,}1834+(0{,}0821)^2\times 0{,}1186+\ldots+(0{,}0854)^2\times 0{,}1001=0{,}015123$$

$$(-0{,}1050)\times(0{,}0027)\times 0{,}1834+(0{,}0821)\times(-0{,}1181)\times 0{,}1186+\ldots+(0{,}0854)\times(-0{,}0513)\times 0{,}1001=0$$

Le tableau donne également la contribution de chaque individu à la formation de l'axe, ou inertie selon cet axe. Cette valeur est définie par :

$$Ctr(L_i,F_k)=\frac{(Masse\ \ L_i)\times(Coord\ \ L_i\ \ selon\ \ F_k)^2}{Valeur\ \ propre\ \ relative\ \ à\ \ F_k}$$

Par exemple, pour Paris et l'axe factoriel N°1 :

$$Ctr(PARIS,F_1)=\frac{0{,}1834\times(-0{,}1050)^2}{0{,}0151}=0{,}1337$$

Ces valeurs sont des contributions relatives (la somme de la colonne vaut 1). On peut donc utiliser des colonnes pour rechercher quels sont les individus-lignes qui ont eu une influence supérieure à la moyenne dans la formation de l'axe factoriel considéré.

Enfin, ce tableau nous donne les cosinus-carrés ou qualités de représentation des individus-lignes par chaque axe factoriel. Ces valeurs sont définies par :

$$QLT(L_i,F_k)=\frac{(Coord\,de\,L_i\,selon\,F_k)^2}{\sum_l (Coord\,de\,L_i\,selon\,F_l)^2}$$

Par exemple :

$$QLT(PARIS,F_1)=\frac{(-0{,}1050)^2}{(-0{,}1050)^2+(0{,}0027)^2+(0{,}1016)^2+(0{,}0107)^2+(0{,}0068)^2+(-0{,}0017)^2+(-0{,}0007)^2}=0{,}5122$$

L'interprétation géométrique de ces valeurs est analogue à celle développée pour l'ACP : c'est le carré du cosinus de l'angle du vecteur représentant "PARIS" dans l'espace à 7 dimensions de sa projection sur le premier axe factoriel.

3.2.3.3 Résultats relatifs aux individus-colonnes

Dans une AFC, les individus-lignes et les individus-colonnes jouent des rôles symétriques. Les résultats relatifs aux individus-colonnes s'interprètent donc de la même façon que les résultats relatifs aux individus-lignes.

Coordonnées Colonne et Contributions à l'Inertie (idf.sta)
Standardisation : Profils ligne et colonne

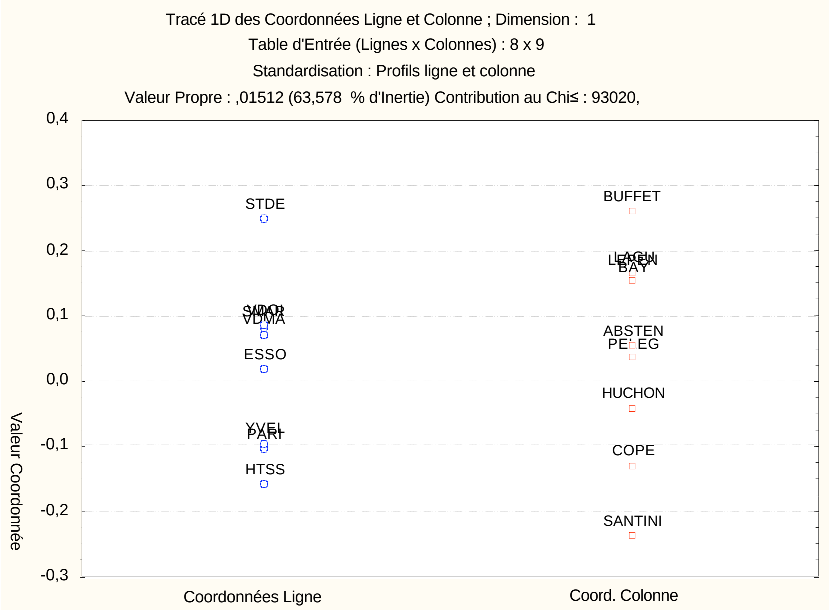
	Colon-ne N°	Coord. Dim.1	Coord. Dim.2	Coord. Dim.3	Masse	Qualité	Inertie Relative	Inertie Dim.1	Cos2 Dim.1	Inertie Dim.2	Cos2 Dim.2	Inertie Dim.3	Cos2 Dim.3
HUCHON	1	-0,0421	-0,0165	0,1024	0,1903	0,9795	0,1024	0,0223	0,1388	0,0112	0,0214	0,5857	0,8194
COPE	2	-0,1305	-0,0513	-0,0089	0,1477	0,9432	0,1299	0,1663	0,8139	0,0833	0,1256	0,0034	0,0038
SANTINI	3	-0,2388	0,0955	-0,0822	0,0960	0,9928	0,2964	0,3622	0,7768	0,1880	0,1241	0,1903	0,0919
LEPEN	4	0,1628	-0,1146	-0,0883	0,0730	0,9912	0,1469	0,1279	0,5539	0,2059	0,2745	0,1670	0,1628
BUFFET	5	0,2581	0,2259	0,0117	0,0429	0,9912	0,2144	0,1890	0,5605	0,4704	0,4295	0,0017	0,0012
LAGU	6	0,1655	-0,0084	-0,0066	0,0238	0,9401	0,0292	0,0431	0,9362	0,0004	0,0024	0,0003	0,0015
PELEG	7	0,0332	-0,0714	-0,0625	0,0149	0,5543	0,0114	0,0011	0,0604	0,0163	0,2796	0,0171	0,2144
BAY	8	0,1514	-0,1198	-0,1211	0,0070	0,9497	0,0161	0,0106	0,4190	0,0216	0,2626	0,0302	0,2681
ABSTEN	9	0,0538	0,0058	-0,0059	0,4044	0,9470	0,0533	0,0775	0,9251	0,0029	0,0106	0,0042	0,0113

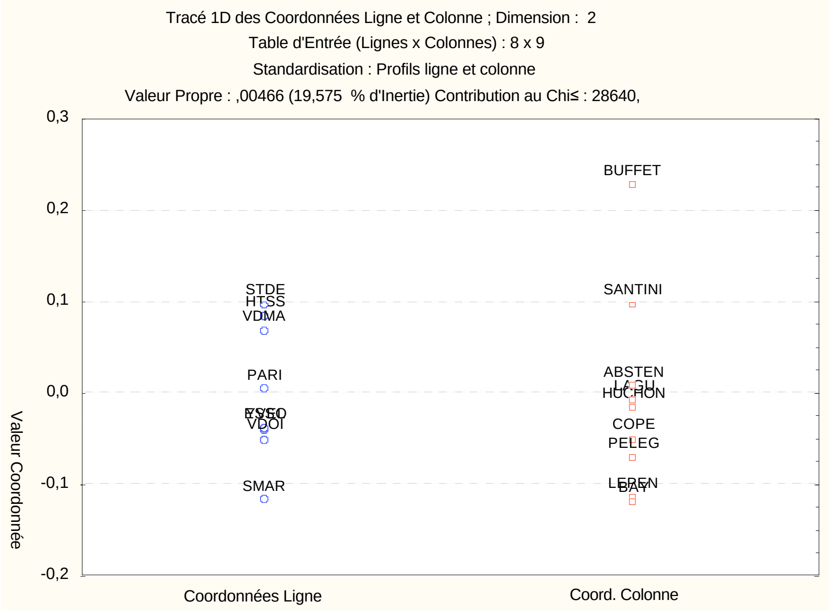
3.2.3.4 Résultats graphiques

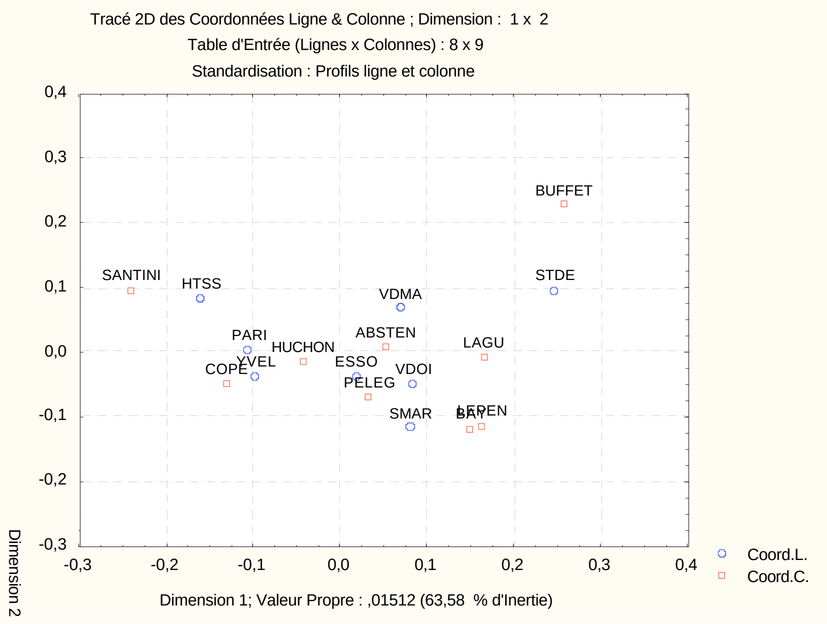
Les transformations et les pondérations introduites rendent tout à fait comparables les valeurs obtenues pour les individus lignes et les individus colonnes. Contrairement à l'ACP, les graphiques factoriels pourront être construits en faisant figurer sur un même graphique les individus lignes et les individus colonnes.

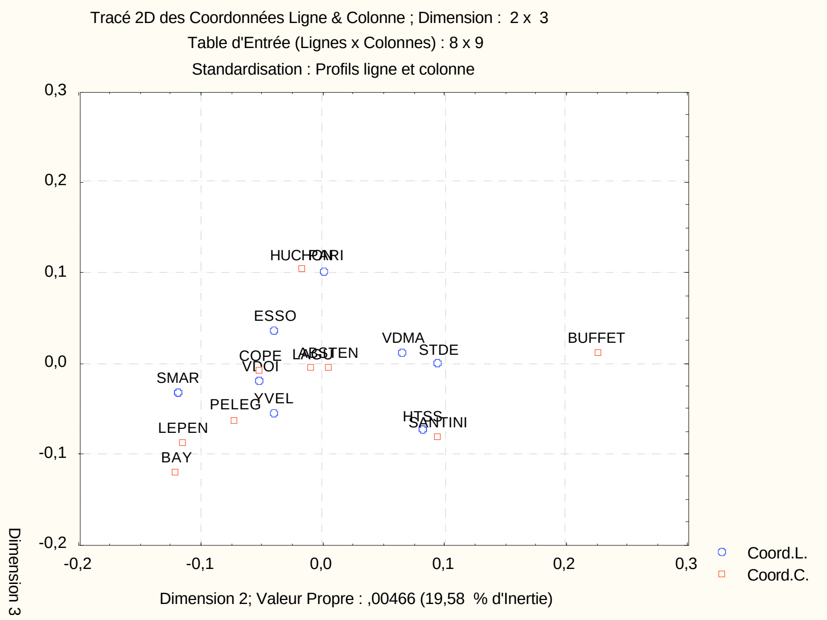
On peut réaliser et essayer d'interpréter des graphiques :

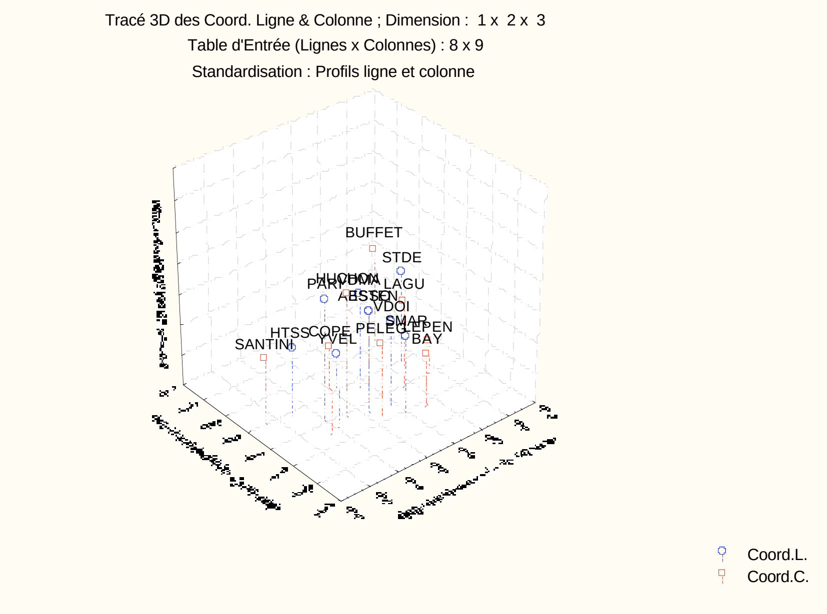
- en dimension 1 : on place les individus le long d'un axe factoriel,
- en dimension 2 : on place les individus dans un plan défini à partir de deux axes factoriels,
- éventuellement, en dimension 3 :on place les individus dans une représentation en perspective d'un espace à 3 dimensions.






Tracé 2D des Coordonnées Ligne & Colonne ; Dimension : 1 x 2
Table d'Entrée (Lignes x Colonnes) : 8 x 9
Standardisation : Profils ligne et colonne
0,4
0,3
0,2
0,1
0,0
-0,1
-0,2
-0,3
-0,3
-0,2
-0,1
0,0
0,1
0,2
0,3
0,4
BUFFET
SANTINI
HTSS
STDE
VDMA
PARI
ABSTEN
HUCHON
LAGU
YVEL
ESSO
COPE
VDOI
PELEG
SMAR
BAY
LEPEN
Dimension 2
Dimension 1; Valeur Propre : ,01512 (63,58 % d'Inertie)
Coord.L.
Coord.C.


Tracé 2D des Coordonnées Ligne & Colonne ; Dimension : 2 x 3
Table d'Entrée (Lignes x Colonnes) : 8 x 9
Standardisation : Profils ligne et colonne
0,3
0,2
0,1
0,0
-0,1
-0,2
-0,2
-0,1
0,0
0,1
0,2
0,3
HUCHON
PARI
ESSO
VDMA
STDE
BUFFET
COPE
LAGU
ABSTEN
VDOI
SMAR
PELEG
YVEL
HTSS
SANTINI
LEPEN
BAY
Dimension 3
Dimension 2; Valeur Propre : ,00466 (19,58 % d'Inertie)
Coord.L.
Coord.C.


Tracé 3D des Coord. Ligne & Colonne ; Dimension : 1 x 2 x 3
Table d'Entrée (Lignes x Colonnes) : 8 x 9
Standardisation : Profils ligne et colonne
BUFFET
STDE
PARI
HUCHON
VDMA
LAGU
ABSTEN
ESSO
VDOI
SMAR
HTSS
COPE
PELEG
LEPEN
SANTINI
YVEL
BAY
Coord.L.
Coord.C.

3.2.3.5 Interprétation géométrique

Les distances entre deux individus-lignes, ou entre un indivivu-ligne et l'origine des axes, peuvent être facilement interprétées. En effet : la distance euclidienne entre deux points-lignes, représentés par leurs coordonnées factorielles est égale à la distance du Φ^2 entre les profils-lignes initiaux.

Par exemple, nous avions vu que :

$$d^2_{\Phi^2}(PARI, SMAR) = \frac{(0{,}2291 - 0{,}1765)^2}{0{,}1903} + \ldots + \frac{(0{,}3847 - 0{,}4133)^2}{0{,}4044} = 0{,}0682$$

Or, le tableau (complet) des scores factoriels des lignes est :

	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5	Facteur 6	Facteur 7
PARI	-0,1050	0,0027	0,1016	0,0107	-0,0068	-0,0017	-0,0007
SMAR	0,0821	-0,1181	-0,0332	0,0231	0,0077	-0,0115	-0,0004
YVEL	-0,0960	-0,0397	-0,0555	0,0029	-0,0015	0,0148	-0,0062
ESSO	0,0183	-0,0393	0,0355	-0,0442	0,0149	-0,0026	-0,0016
HTSS	-0,1586	0,0824	-0,0752	-0,0042	-0,0011	-0,0104	0,0019
STDE	0,2478	0,0954	-0,0017	0,0006	-0,0096	-0,0026	-0,0070
VDMA	0,0706	0,0667	0,0115	0,0151	0,0208	0,0100	0,0066
VDOI	0,0854	-0,0513	-0,0206	-0,0134	-0,0231	0,0050	0,0092

On vérifie que :

$$d^2_{eucl}(PARI', SMAR') = (-0{,}1050 - 0{,}0821)^2 + \ldots + (-0{,}0007 + 0{,}0004)^2 = 0{,}0682$$

De même, on avait établi que :

$$d^2_{\Phi^2}(PARI, Moyenne) = \frac{(0{,}2291 - 0{,}1903)^2}{0{,}1903} + \ldots + \frac{(0{,}3847 - 0{,}4044)^2}{0{,}4044} = 0{,}0215$$

Et l'on a :

$$d^2_{eucl}(PARI', O) = (-0{,}1050)^2 + \ldots + (-0{,}0007)^2 = 0{,}0215$$

La même propriété s'applique aux colonnes. Le tableau complet des scores factoriels des colonnes est donné par :

	Facteur 1	Facteur 2	Facteur 3	Facteur 4	Facteur 5	Facteur 6	Facteur 7
HUCHON	-0,0421	-0,0165	0,1024	-0,0157	0,0024	-0,0023	-0,0020
COPE	-0,1305	-0,0513	-0,0089	0,0325	0,0108	-0,0038	0,0013
SANTINI	-0,2388	0,0955	-0,0822	-0,0225	-0,0035	-0,0032	-0,0009
LEPEN	0,1628	-0,1146	-0,0883	-0,0174	0,0017	-0,0101	-0,0034
BUFFET	0,2581	0,2259	0,0117	0,0178	0,0259	-0,0069	-0,0027
LAGU	0,1655	-0,0084	-0,0066	-0,0212	-0,0020	-0,0297	0,0204
PELEG	0,0332	-0,0714	-0,0625	-0,0499	0,0601	0,0423	0,0148
BAY	0,1514	-0,1198	-0,1211	-0,0160	0,0350	-0,0014	-0,0356
ABSTEN	0,0538	0,0058	-0,0059	0,0055	-0,0100	0,0060	0,0004

On avait établi que :

$$d^2_{\Phi^2}(BUFFET, SANTINI) = \frac{(0{,}1480 - 0{,}1934)^2}{0{,}1834} + \ldots + \frac{(0{,}0936 - 0{,}0811)^2}{0{,}1001} = 0{,}2753$$

On retrouve ici :

$$d^2_{eucl}(BUFFET', SANTINI') = (0{,}2581 + 0{,}2388)^2 + \ldots + (-0{,}0027 + 0{,}0009)^2 = 0{,}2753$$

La proximité entre un point-ligne L et un point-colonne C ne possède pas d'interprétation géométrique immédiate. En revanche, l'angle de sommet O dont les côtés passent par L et C a la propriété suivante :

- si l'angle (OL, OC) est aigu, la modalité-ligne L et la modalité colonne C s'attirent (taux de liaison positif)

- si l'angle (OL, OC) est aigu, la modalité-ligne L et la modalité colonne C se repoussent (taux de liaison négatif)
- si l'angle (OL, OC) est droit, la modalité-ligne L et la modalité colonne C n'interagissent pas (taux de liaison voisin de 0).

3.2.3.6 Reconstitution des données

Il est possible de reconstituer les données à partir des scores factoriels des lignes et des colonnes. En effet, on peut montrer la relation suivante entre les taux de liaison t_{ij}, les scores factoriels des lignes, les scores factoriels des colonnes et les valeurs propres :

$$t_{ij} = \sum_{Axes\ \ factoriels} \frac{(Score\ \ fact.\ ligne\ \ i\ \ selon\ \ axe\ \ \alpha)(Score\ \ fact.\ colonne\ \ j\ \ selon\ \ axe\ \ \alpha)}{\sqrt{Valeur\ \ propre\ \ associée\ \ à\ \ l'axe\ \ \alpha}}$$

Par exemple, le taux de liaison entre PARI et la liste HUCHON peut être retrouvé à l'aide du calcul suivant :

$$t_{11} = \frac{(-0{,}1050)(-0{,}0421)}{\sqrt{0{,}015123}} + \frac{(0{,}0027)(-0{,}0165)}{\sqrt{0{,}004656}} + \ldots + \frac{(-0{,}0007)(-0{,}0020)}{\sqrt{0{,}000024}} = 0{,}20406$$

Connaissant les profils moyens des lignes et des colonnes, et l'effectif total N, l'ensemble des données peut ainsi être retrouvé.