

INTRODUCTION À L'ANALYSE FACTORIELLE DES CORRESPONDANCES

Dominique LAFFLY

Maître de Conférences, Université de Pau
Laboratoire Société Environnement Territoire

UMR 5603 du CNRS et Université de Pau
Domaine Universitaire, IRSAM, 64000 PAU
Tél : 05 59 92 31 23 Fax : 05 59 80 83 39
Mail : dominique.laffly@univ-pau.fr

L'analyse factorielle des correspondances vise à rassembler en un nombre réduit de dimensions la plus grande partie de l'information initiale en s'attachant non pas aux valeurs absolues mais aux correspondances entre les variables, c'est-à-dire aux valeurs relatives. Cette réduction est d'autant plus utile que le nombre de dimensions initial est élevé. La notion de "réduction" est commune à toutes les techniques factorielles – c'est-à-dire où l'on extrait des facteurs – l'AFC offre la particularité (contrairement aux ACP) de fournir un espace de représentation commun aux variables et aux individus. Pour cela l'AFC raisonne à partir de tableau réduit ou de fréquences. Prenons comme exemple la matrice de contingence suivante :

	J	AF	AH	V	Mains		J	AF	AH	V	Mains	
Ch	69	172	133	27	401	réduction $f_{i,j} = k_{i,j}/k_{..}$	Ch	0.069	0.172	0.133	0.027	0.401
Ja	41	84	118	11	254		Ja	0.041	0.084	0.118	0.011	0.254
Cl	18	127	157	43	345		Cl	0.018	0.127	0.157	0.043	0.345
Pieds	128	383	408	81	1000		Pieds	0.128	0.383	0.408	0.081	1

Où : J=jeunes ; AF=adulte femme ; AH=adulte homme ; V=vieux ; Ch=chansons ; Ja=jazz ; Cl=classique

$f_{i,j}$ fréquence de la case i, j ;

$k_{i,j}$ valeur brute de la case i, j ;

$k_{..}$ somme de la matrice initiale (1000).

Supposons que l'on désire observer la valeur relative des réponses par rapport aux colonnes, c'est-à-dire comment se ventilent les choix des jeunes selon le type de musique, par exemple. Pour ce on transforme le tableau réduit en pondérant chaque case par les pieds :

	J	AF	AH	V	Mains
Ch	0.539	0.449	0.325	0.333	1.646
Ja	0.32	0.219	0.289	0.135	0.963
Cl	0.14	0.331	0.384	0.53	1.385
Pieds	1	1	1	1	4

On pourrait dès lors en faire une représentation graphique dans un espace à trois dimensions (classique, jazz et chanson). Il est même possible d'associer à tous les points un cercle dont la surface est proportionnelle à la valeur des mains (figure A8).

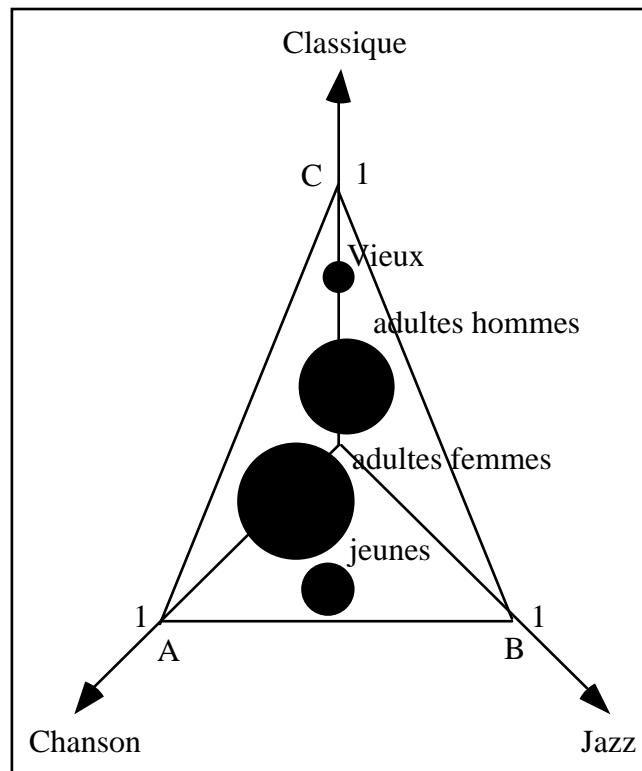


Figure A8 : Projection des variables selon les individus

Tous les points se situent dans le plan défini par le triangle de coordonnées 1, 1, 1. Lorsque cette condition est vérifiée on se trouve dans le cas d'une représentation duale. Supposons désormais que l'on désire observer la valeur relative des réponses par rapport aux lignes, c'est-à-dire comment se ventilent les types de musique selon l'âge et le sexe des individus. Pour ce on transforme le tableau réduit en pondérant chaque case par les mains :

	J	AF	AH	V	Mains
Ch	0.172	0.429	0.331	0.067	1
Ja.	0.161	0.33	0.464	0.043	1
Cl	0.052	0.368	0.455	0.124	1
Pieds	0.385	1.127	1.25	0.234	3

Cela revient à considérer un nuage de trois points (classique, jazz et chanson) dans un espace à quatre dimensions (jeunes, adultes femmes, adultes hommes et vieux). Comme il n'y a que trois points ils sont forcément dans un même plan. En calant ce dernier avec le plan précédent de manière à faire coïncider les axes on obtiendrait à peu près la même représentation, en tous cas l'origine des axes serait la même. Les trois points Cl, Ja et Ch ne se superposeraient pas A, B et C mais ils s'aligneraient sur les mêmes directions à une échelle différente. On aurait le choix alors soit de représenter J, AF, AM et V à l'intérieur de A, B et C soit de faire l'inverse puisque les origines coïncident. L'AFC permet de résoudre le problème d'échelle et d'obtenir le même représentation. Pour cela on introduit une pondération. L'AFC consiste, dans l'une des analyses, à pondérer les pieds par la racine carrée des mains, et dans l'autre analyse, de pondérer les mains par la racine carrée de pieds.

Dans la première analyse $f_{i,j}/f_j$ devient $f_{i,j}/f_j * \sqrt{f_i}$, dans la deuxième analyse $f_{i,j}/f_i$ devient $f_{i,j}/f_i * \sqrt{f_j}$ avec : SQRT pour racine carrée, $f_{i,j}$ fréquence de la case i, j ; f_j somme marginale en colonnes (pieds) et f_i somme marginale en lignes (mains). Dans ce qui suit la notation $f_{..}$ renvoie à la somme de toutes les fréquences. Soit, pour la première analyse, la matrice suivante :

	J	AF	AH	V
Ch	0.85	0.709	0.514	0.526
Ja.	0.63	0.435	0.573	0.269
Cl	0.24	0.564	0.655	0.9

La première conséquence de cette pondération est une déformation du triangle ABC initial, les trois sommets ne sont plus situés à la même distance de l'origine mais à des distances différentes dépendantes de la pondération. De plus, la perpendiculaire abaissée sur le triangle de l'origine 0 se trouve au centre de gravité G des quatre points J, AF, AM et V. Donc, l'axe OG est un axe d'inertie du nuage (axe trivial). Enfin, les moments d'inertie des deux nuages issus des deux analyses sont égaux entre eux, leur valeur commune n'est autre que le lien ou information mutuelle entre deux caractères. Contrairement au Chi^2 de Pearson, le **lien** exprime la liaison entre deux variables indépendamment du nombre d'individus de la population. Le lien est évidemment nul pour les tableaux de distributions théoriques. Le lien

est le rapport de la valeur du $Chi2$ par la population. Il est plus commode de procéder au calcul du lien sur des tableaux réduits, la formule devient alors :

$$Lien = \sum_{i=1}^n \sum_{j=1}^m \frac{(f_{i,j} - f_{..})^2}{f_{..}}$$

Le lien évolue de 0 (indépendance totale) à 1 (dépendance totale) alors que le $Chi2$ varie réciproquement de 0 à n (nombre d'individus dans la population). Le calcul du lien offre une particularité importante pour l'AFC. D'autre part le calcul du lien peut se simplifier :

Sachant que : $\sum f_j = 1$ et $\sum f_i = 1$

$$\sum f_{..} = 1$$

$$\sum f_{ij} = 1$$

On a alors :

$$Lien = \sum_{i=1}^n \sum_{j=1}^m \frac{(f_{i,j})^2}{f_{..}} - 1$$

Nous avons dit plus haut qu'il existait une signification forte entre l'inertie du nuage de points de l'AFC et le lien. En effet, le moment d'inertie d'un point de masse m et coordonnées x , y et z par rapport au point O origine est égal à $m*(x^2+y^2+z^2)$. La coordonnée d'un point étant ici de $f_{i,j}/f_j * \sqrt{f_i}$ son carré devient $f_{i,j}^2/f_j^2*f_i$ et en le multipliant ensuite par sa masse qui est f_j on obtient $f_{i,j}^2/f_j*f_i$. Cette expression est parfaitement symétrique par rapport aux sommes marginales et on obtient le même résultat pour les deux analyses lorsque l'on somme tous les calculs. Cette somme (XX) n'est autre que le lien ou le premier terme est le moment d'inertie du nuage par rapport à O et le second terme -1 le transfert de ce moment au centre de gravité du nuage. Du fait que $\sum f_j = 1$ et $\sum f_i = 1$ on se rend compte que :

- la masse de chaque nuage est égale à 1 ;
- les coordonnées du centre de gravité G du nuage sont égales à $\sqrt{f_i}$ dans une analyse et $\sqrt{f_j}$ dans l'autre ;
- la distance OG est égale à 1 dans les deux analyses.

On peut donc calculer directement le moment d'inertie de chaque nuage par rapport à son centre d'inertie soit : $\sum [(f_{i,j}) - (f_{..})]^2 / (f_{..})$ (b). Les deux expressions (a) et (b) conduisent à un

résultat identique, la seconde présente l'avantage de faire le calcul case par case. La valeur de chaque case n'étant autre ici que la contribution au lien. Cette identité des moments d'inertie élémentaires avec les composantes du lien est une caractéristique essentielle de l'AFC. C'est pourquoi on dit souvent que l'AFC suit une métrique du χ^2 du fait de la relation étroite $\chi^2 = \text{lien} * n$.

Nous le verrons plus loin, du fait que l'analyse est la même pour les deux transformations du nuage on dit que les calculs sont duaux l'un de l'autre. On peut donc n'en choisir qu'un, le plus facile c'est-à-dire en formant la matrice V de dimensions (k, k) plutôt que la matrice W de dimensions (n, n) en supposant toujours $(n \geq k)$. Si la matrice est haute on aura $V=X'X$ et si la matrice est basse $V=XX'$ (où X' est la matrice transposée de X).

On dresse alors les deux matrices suivantes :

	J	AF	AH	V		J	AF	AH	V
Ch	0.218	0.076	-0.118	-0.107	Ch	0.078	0.047	-0.076	-0.03
Ja	0.132	-0.069	0.070	-0.235	Ja	0.047	-0.043	0.045	-0.067
Cl	-0.348	-0.023	0.068	0.316	Cl	-0.124	-0.014	0.043	0.09

terme courant $(f_{i,j} - (f_{..})) / f_{.j} * \sqrt{f_{i.}}$ terme courant $(f_{i,j} - (f_{..})) / \sqrt{f_{..}}$

La première matrice est celle des coordonnées selon la pondération rappelée plus haut. Elle servira plus tard à calculer les coordonnées des points sur les axes factoriels. La seconde matrice sera multipliée par sa transposée afin d'obtenir la matrice d'inertie (ou de variance-covariance) qui sera diagonalisée pour obtenir les valeurs propres du nuage de points. Soit la matrice suivante :

	Ch	Ja	Cl
Ch	0.0149	0.000235	-0.0164
Ja	0.000235	0.0105	-0.0093
Cl	-0.0164	-0.0093	0.0257

C'est une matrice symétrique dont la trace (somme des termes de la diagonale, c'est également la somme de tous les éléments de la matrice V élevés au carré) est égale à 0.0511. Dans une telle matrice les éléments diagonaux sont des éléments de variance (le carré d'éléments de V) les autres éléments sont dits de covariance (produit croisé de termes de V). Une telle matrice fait correspondre à un point Y de coordonnées y_1, y_2 et y_3 un point X de coordonnées x_1, x_2 et x_3 selon la relation suivante :

$$y_1 = 0.0149x_1 + 0.000235x_2 - 0.0164x_3$$

$$y_2 = 0.000235x_1 + 0.0105x_2 - 0.0093x_3$$

$$y_3 = -0.0164x_1 - 0.0093x_2 + 0.0257x_3$$

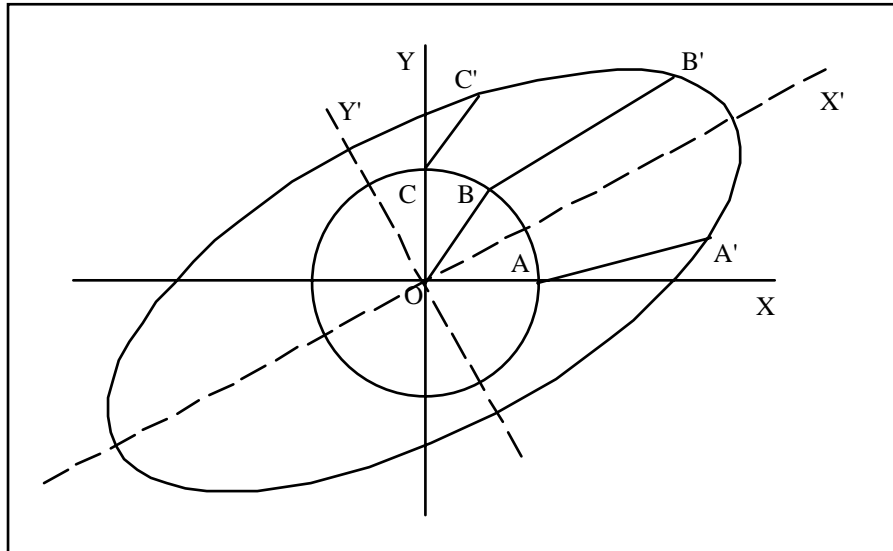


Figure A9: Transformation par la matrice d'inertie

C'est-à-dire que la matrice transforme la figure dessinée par les points de X en une autre figure dessinée par les points de Y . Si on avait un cercle (deux dimensions) la transformation donnerait une ellipse. Du fait de cette transformation la matrice d'inertie est considérée comme un ellipsoïde d'inertie. Cette dernière peut être assimilée à la projection du nuage de points qui nous intéresse (figure A9). Les directions de plus grand allongement de l'ellipse représentent les tendances dominantes, il est intéressant de les déterminer, ce sont les axes de symétrie ou axes principaux. Nous empruntons à la mécanique la notion d'inertie. Il y a identité absolue entre moyenne arithmétique (statistique) et centre de gravité (mécanique) – variables de position – ainsi qu'entre variance (statistique) et inertie (mécanique). En mécanique la notion d'inertie joue sur les corps tournants. Un corps a d'autant plus d'inertie autour d'un axe qu'il faut lui appliquer d'énergie pour le mettre en rotation autour d'un axe. De plus, pour la même quantité d'énergie, en faisant passer l'axe de rotation sur l'ensemble de la surface du corps (imaginons un patatoïde en deux dimensions) la vitesse de rotation passera par un maximum et un minimum. Les positions respectives de l'axe correspondront au moment de plus petite et de plus grande inertie, ce seront également les positions des axes d'inertie ou axes principaux.

Ce qui intervient dans l'inertie c'est la masse et le carré de la distance au centre de gravité : $I = MD^2$. En remplaçant la masse par la fréquence et la distance par écart l'inertie devient la variance élémentaire. De plus, tout corps solide aussi biscornu soit-il se comporte de façon identique à un ellipsoïde d'inertie à trois axes de symétrie :

- grand axe : inertie minimale (tous les points sont proches) ;
- petit axe : inertie maximale (tous les points sont éloignés) ;
- axe moyen : situation intermédiaire.

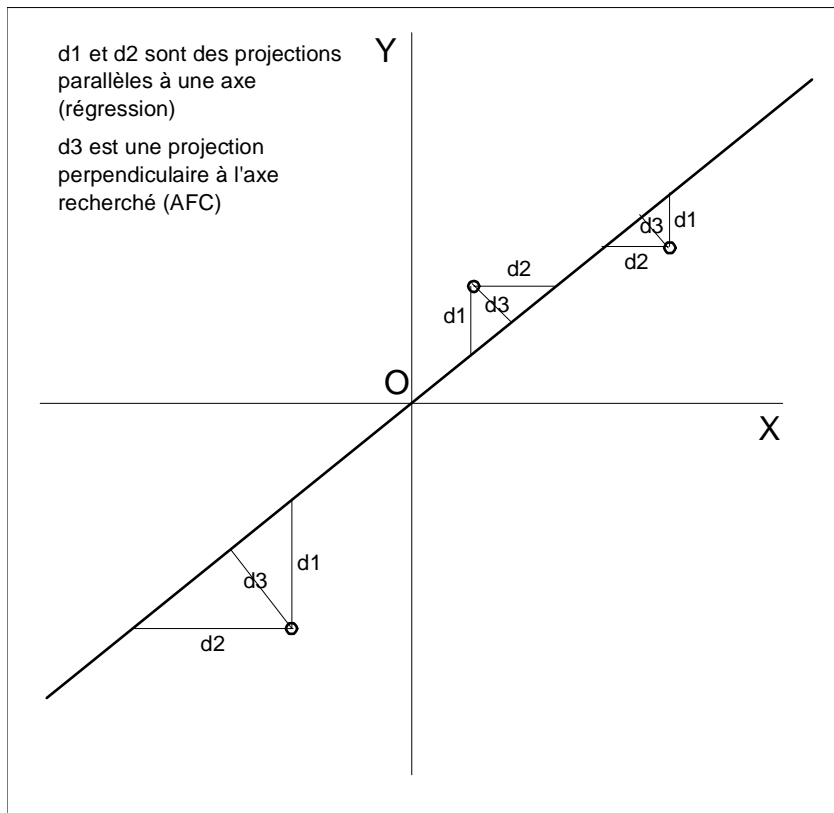


Figure 10 : Différentes méthodes de projection

À propos des différents axes on insiste sur la distance des points à ceux-ci on pose donc le problème de la projection des points sur un axe. Le principe consiste à trouver les situations pour lesquelles la somme des distances au carré projetée est soit minimale, soit maximale soit intermédiaire. Il existe différentes projections (figure A10), en analyse des données nous utilisons une projection perpendiculaire à l'axe, contrairement aux méthodes de régression linéaire, selon une méthode appelée ajustement par les moindres carrés. La particularité de l'AFC repose sur la distance utilisée, celle-ci n'est pas euclidienne mais celle du *Chi2* comme nous le verrons plus bas.

Trouver les axes principaux d'une matrice revient à diagonaliser la matrice, elle sera alors de la forme :

	Ch	Ja	Cl
Ch	λ_1	0	0
Ja	0	λ_2	0
Cl	0	0	λ_3

Avec : $\lambda_1 + \lambda_2 + \lambda_3 = \text{trace} = 0.0511$

λ_1, λ_2 et λ_3 les valeurs propres du nuage

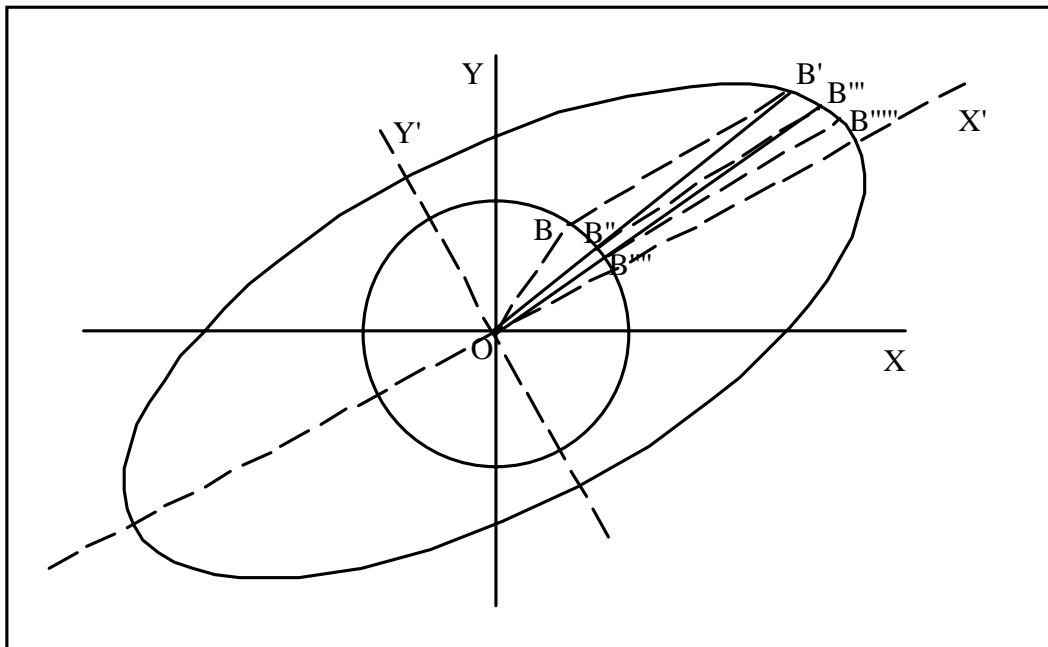


Figure A11 : Diagonalisation d'une matrice

La trace est inchangée donc la diagonale ne change pas la variance mais on supprime les termes de covariances. La diagonalisation d'une matrice est une opération relativement difficile à mettre en œuvre. Il existe d'ailleurs différentes méthodes d'ailleurs. Nous présenterons celles dite de la dragée qui a le grand avantage de pouvoir se comprendre facilement graphiquement (figure A11). Avec cette méthode on cherche à tendre asymptotiquement vers la direction des axes d'inertie. À partir d'un point quelconque (ici B) la première transformation donne le point B' . On raccourcit OB' à la dimension du rayon du cercle, on trouve le point B'' qui est déjà plus proche du grand axe. Ce point B'' subit une nouvelle transformation par la matrice d'inertie ce qui donne B''' . Ce dernier est réduit au

rayon du cercle en B'''' et ainsi de suite jusqu'à ce que la réduction se situe sur l'axe d'inertie. La valeur λI correspond au cumul des carrés des distances des transformations des points. Notons que ce raisonnement s'étend facilement à n dimensions.

La diagonalisation de la matrice issue de la seconde analyse donnerait les mêmes valeurs propres c'est ce qu'on appelle alors un espace dual. Concrètement cela se traduit par la caractéristique de double représentativité propre à l'AFC c'est-à-dire que la projection des individus ou des caractères se fait dans un espace identique, les proximités entre points auront donc une forte signification thématique comme nous le verrons plus loin. Notons enfin à propos de l'inertie que : inertie totale = inertie expliquée par l'axe + inertie autour de l'axe et que les valeurs propres de la matrice d'inertie sont les inerties expliquées, leur somme est égale à la trace elle-même égale à l'inertie totale.

Dans notre exemple la diagonalisation de la matrice nous donne :

	Ch	Ja	Cl
Ch	0.0396	0	0
Ja	0	0.0115	0
Cl	0	0	0

La troisième valeur propre est égale à 0 parce que les quatre points J , AM , AF et V sont dans un même plan. L'axe d'inertie de ce dernier, nous l'avons dit plus haut, correspond à la droite perpendiculaire à ce plan. Sa valeur est de 1, il est sans intérêt pour l'analyse puisqu'il n'explique aucune part de la variance, on l'appelle **axe trivial**. En revanche, on retrouve bien $\lambda_1 + \lambda_2 = 0.0511 = \text{trace} = \text{lien}$.

Le premier axe explique $0.0396/0.0511 = 77.5\%$ de la variance (ou de l'inertie) et le second $0.0115/0.0511 = 22.5\%$. Outre ce calcul, les valeurs propres permettent de déterminer les coordonnées des points dans l'espace factoriel. Il s'agit d'une projection qui s'obtient par un produit scalaire pour chaque point et pour chaque axe :

$$Jeune_{S_{Axe1}} = (0.218 * 0.0396) + (0.132 * 0) + (-0.348 * 0) = 0.0086328$$

$$Jeune_{S_{Axe2}} = (0.218 * 0) + (0.132 * 0.0115) + (-0.348 * 0) = 0.001518$$

...

D'un point de vue matricielle cela revient à multiplier la transposée de la matrice de coordonnées par la matrice diagonalisée :

	Ch	Ja	Cl		Ch	Ja	Cl		Axe1	Axe2
J	0.218	0.132	-0.348	Ch	0.0396	0	0	J	0.0086328	0.001518
AF	0.076	-0.069	-0.023	* Ja	0	0.0115	0	= AF	0.0030096	-0.0007935
AH	-0.118	0.070	0.068	Cl	0	0	0	AM	-0.0046728	0.00805
V	-0.107	-0.235	0.316					V	-0.0042372	-0.0027025

La première analyse est alors terminée. Nous l'avons dit plus haut nous nous trouvons dans un cas de relations duales, il est inutile de faire la seconde analyse, les coordonnées des 3 points "musique" se déduisent immédiatement de celles des quatre points "population" par les relations :

$$i_{1axe1} = (1/K_{i1} * \sqrt{\lambda 1}) * (K_{1,1}V_{1axe1} + K_{1,2}V_{2axe1} \dots K_{1,n}V_{naxe1})$$

$$i_{1axe2} = (1/K_{i1} * \sqrt{\lambda 2}) * (K_{1,1}V_{1axe2} + K_{1,2}V_{2axe2} \dots K_{1,n}V_{naxe2})$$

...

$$i_{1axe k} = (1/K_{i1} * \sqrt{\lambda k}) * (K_{1,1}V_{1axe k} + K_{1,2}V_{2axe k} \dots K_{1,n}V_{naxe k})$$

...

...

...

$$i_{paxe1} = (1/K_{ip} * \sqrt{\lambda 1}) * (K_{p,1}V_{1axe1} + K_{p,2}V_{2axe1} \dots K_{p,n}V_{naxe1})$$

$$i_{paxe2} = (1/K_{ip} * \sqrt{\lambda 2}) * (K_{p,1}V_{1axe2} + K_{p,2}V_{2axe2} \dots K_{p,n}V_{naxe2})$$

...

$$i_{paxe k} = (1/K_{ip} * \sqrt{\lambda k}) * (K_{p,1}V_{1axe k} + K_{p,2}V_{2axe k} \dots K_{p,n}V_{naxe k})$$

où : $i_1, i_2 \dots i_p$ sont les p individus de la **série brute de départ** ;

$axe_1, axe_2 \dots axe_k$ sont les k axes d'inertie ;

$\sqrt{\lambda 1}, \sqrt{\lambda 2} \dots \sqrt{\lambda k}$ sont k valeurs propres ;

K_{ij} sont les données brutes de la séries ;

$V_{naxe k}$ sont les coordonnées des n variables sur les k axes d'inertie.

Par exemple cela donnerait pour le point Ch sur l'axe 1 :

$$Ch_{axe1} = 1/(401 * \sqrt{0.0396}) * ((69 * 0.0086328) + (172 * 0.0030096) - (133 * 0.0046728) - (27 * 0.0042372))$$

$$Ch_{axe1} = 1/(101 * 0.19899) * (0.5956632 + 0.5176512 - 0.6214824 - 0.11449044)$$

$$Ch_{axe1} = 0.049756219 * (0.5956632 + 0.5176512 - 0.6214824 - 0.11449044)$$

$$Ch_{axe1} = 0.018775089$$

En langage savant la formule ci-dessus s'énonce en disant « qu'à une affinité de rapport $1/\sqrt{\lambda}$ près, les coordonnées des points du nuage dual sur un axe factoriel sont les coordonnées du barycentre du nuage direct, pondérées par l'importance relative de la rubrique duale dans chacun d'entre eux ».

Tous les calculs détaillés plus haut sont totalement transparents pour l'utilisateur, de plus ils sont extrêmement rapides sur les ordinateurs de la nouvelle génération. Notons cependant que vue la précision des résultats il arrive que, selon le programme utilisé, on observe des inversions de coordonnées. Autrement dit ce qui était positif devient négatif et inversement. Cela n'est d'aucune importance pour la suite puisque les interprétations en matière d'AFC repose sur les positions relatives des points les unes par rapport aux autres.

Graphiquement, les résultats de l'analyse donneraient le dessin en haut à gauche de la figure suivante. Le schéma situé en bas à droite de la figure met en avant une caractéristique importante de l'AFC, l'équivalence distributionnelle. Imaginons que pour les besoins de l'étude on ait besoin également de distinguer le sexe des catégories « vieux » et « jeunes » on obtiendrait alors 6 classes de populations. Si les proportions entre JF et JM se trouvaient être identique de celles des « Jeunes » à l'origine les deux nouvelles catégories se trouveraient situés exactement au même point dans l'espace factoriel, celui des « jeunes » de la première analyse. C'est une propriété importante car elle donne à l'AFC une grande indépendance par rapports aux nomenclatures en permettant sans difficulté la réunion de plusieurs classes dont le profil est identique ou voisin. Ce n'est évidemment pas souvent le cas alors, pourvu qu'une classe n'ait pas un poids trop élevé par rapport à l'ensemble, sa décomposition s'opère en sous-classes dont les points représentatifs ont approximativement l'ancien point comme centre de gravité.

Remarquons dans cet exemple que la quantité totale d'information a augmentée en créant de nouvelles classes, elle est passée de 0.0511 à 0.058. C'est une conséquence du théorème de Huygens : quand on décompose une masse en deux sous masses, sans changer le centre de gravité, on ajoute à l'inertie totale l'inertie propre de l'haltère ainsi créé. Une autre conséquence sera que plus le nombre de classes est élevée moins l'inertie des axes sera importante, ce qui ne veut pas dire qu'il n'épuise pas correctement l'inertie totale. Mieux vaut donc interpréter les valeurs propres et leur taux de décroissance dans ces cas là.

Avant de conclure cette présentation de l'AFC notons que d'autres informations sont fournies par les calculs. On connaît déjà les valeurs propres, l'inertie des axes, la masse des variables et les coordonnées sur les axes. On calcule également la contribution des points pour les différents axes ainsi que le cosinus de l'angle des points par rapport aux axes. La contribution (souvent exprimée en pour mille) indique, comme son nom l'indique, le rôle joué par un point dans la détermination de l'axe. C'est particulièrement utile pour l'interprétation thématique, cependant n'oublions jamais qu'un poids de forte masse sera « attiré » par le centre de gravité du nuage sa contribution calculée sera réduite par cet effet et, inversement, un point de faible masse mais éloigné de G aura une contribution exagérée. Ceci tient au calcul de l'inertie qui repose sur le produit de la masse par le carré de la distance au point G. Le cosinus de l'angle (qui varie entre -1 et $+1$) permet d'interpréter la perspective qu'on ne visualise pas dans une représentation en deux dimensions. Lorsque le cosinus sera faible on peut considérer que le point est proche de l'axe, qu'il joue donc beaucoup dans sa définition et, inversement, plus il s'en éloigne moins son rôle sera prédominant dans l'interprétation.

La figure A12 présente graphiquement les résultats de l'AFC. Les points variables et individus sont projetés en fonction de leurs coordonnées sur un graphe de type XY (deux dimensions) ou XYZ (trois dimensions). Généralement on interprète des graphes 2D réalisés à partir des 3 premiers axes, ceux qui épuisent majoritairement l'inertie. Avec 3 axes on dessine alors 3 plans 1/2, 1/3 et 2/3 à partir desquels il faut tenter de reconstruire intellectuellement les trois dimensions du nuage étudiés. Certains logiciels permettent une vue 3D interactive intéressante pour visualiser la profondeur du nuage mais limitée pour une interprétation détaillée.

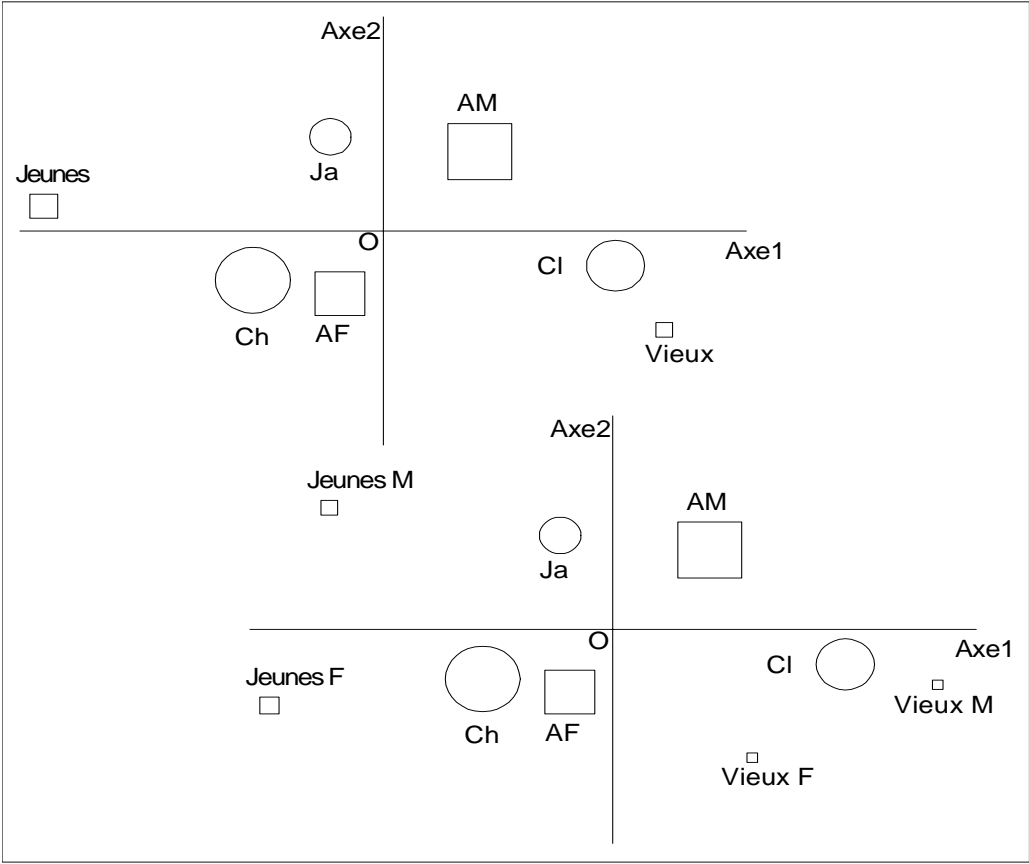


Figure A12 : Représentation graphique des résultats de l'AFC