

Analyse des données avancée

Chapitre 10 : Profils, barycentres et lecture conjointe



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Les relations barycentriques

Le plan factoriel

Les aides à l'interprétation en AFC

Ce qu'il faut retenir

Les relations barycentriques

La transition en AFC

$$F_k(i) = \frac{1}{\sqrt{\lambda_k}} \sum_j \frac{f_{ij}}{f_{i.}} G_k(j)$$

$$G_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_i \frac{f_{ij}}{f_{.j}} F_k(i)$$

Ce qu'elles disent

Une ligne se place au barycentre des colonnes, pondéré par son profil. L'étudiant vient à 63,6 % pour le prix : il est tiré vers *Prix* à hauteur de 63,6 %.

Une colonne se place au barycentre des lignes, selon le même principe. Chaque nuage est le barycentre de l'autre : c'est parfaitement symétrique, et c'est ce qui n'existait pas en ACP.

Pourquoi $1/\sqrt{\lambda_k}$

Le barycentre d'un nuage est toujours **à l'intérieur** de ce nuage. Sans correction, les lignes seraient donc plus resserrées que les colonnes — et les colonnes plus resserrées que les lignes. C'est contradictoire.

Le facteur $1/\sqrt{\lambda_k} \geq 1$ **dilate** chaque nuage pour que les deux tiennent à la même échelle.

La vérification, sur l'axe 1

$\lambda_1 = 0,2267$, donc $1/\sqrt{\lambda_1} = 2,100$.

Barycentre exact de l'étudiant :

$$0,636 \times (-0,544) + 0,091 \times 0,523 + 0,227 \times (-0,366) + 0,045 \times 0,419 = -0,363$$

Après dilatation : $-0,363 \times 2,100 = -\mathbf{0,762}$ — la coordonnée du tableau.

Le résultat que l'ACP n'avait pas

La superposition est ici légitime

En ACP, le chapitre 6 interdisait de lire une distance entre un individu et une variable : les deux nuages vivaient dans des espaces différents.

En AFC, chaque nuage est **littéralement** le barycentre de l'autre. Les deux se représentent sur le même graphique, et **la proximité entre une ligne et une colonne se commente**.

C'est la seule méthode du cours où cette lecture soit permise — et c'est ce qui fait de l'AFC l'outil préféré des études qualitatives.

Une précaution subsiste

Une ligne proche d'une colonne signifie que cette ligne **sur-consomme** cette colonne par rapport à l'indépendance.

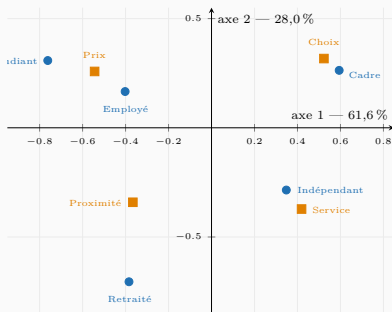
Elle ne signifie pas que c'est sa modalité majoritaire. Le retraité est proche de *Proximité* parce qu'il la cite deux fois et demie plus que la moyenne — 56,7 contre 22,7 % — non parce que c'est un gros effectif : ils ne sont que dix-sept.

Le plan factoriel

Les cinq catégories

Ligne	F_1	F_2	ctr 1	ctr 2	$\cos^2$ plan
Étudiant	-0,762	0,308	18,8	6,8	0,95
Employé	-0,402	0,166	24,0	9,1	1,00
Cadre	0,594	0,263	37,3	16,1	0,93
Indépendant	0,348	-0,285	13,4	19,8	0,76
Retraité	-0,384	-0,705	6,5	48,3	0,84

Exemple					
Colonne	F_1	F_2	ctr 1	ctr 2	$\cos^2$ plan
Prix	-0,544	0,258	36,5	18,2	0,94
Choix	0,523	0,317	33,0	26,7	0,95
Proximité	-0,366	-0,341	13,4	25,5	0,79
Service	0,419	-0,372	17,1	29,6	0,85



Le privilège de l'AFC : les deux nuages sur le même plan.

Lignes et colonnes sont projetées dans le même repère — ce qu'aucune ACP ne permet.

Axe 1 : *Prix* et les étudiants et employés d'un côté ; *Choix*, *Service* et les cadres et indépendants de l'autre. C'est la contrainte budgétaire.

Axe 2 : *Proximité* et les retraités, seuls en bas.

Attention : la proximité ne s'interprète que d'un nuage à l'autre. « Retraité est près de Proximité » se lit ; « Retraité est près d'Employé » ne se lit pas de la même façon.

Axe 1 : la contrainte budgétaire

D'un côté **Prix**, avec les étudiants et les employés. De l'autre **Choix** et **Service**, avec les cadres et les indépendants.

Les quatre contributions dominantes sont *Prix* (36,5 %) et *Choix* (33,0 %) pour les colonnes, les cadres (37,3 %) et les employés (24,0 %) pour les lignes. **L'axe oppose ceux qui viennent parce que c'est moins cher à ceux qui viennent parce qu'il y a plus de choix.**

Axe 2 : les retraités, seuls

Le second axe est fabriqué à 48,3 % par les retraités, contre *Proximité* et *Service* au sud.

Un axe porté à moitié par une seule modalité de 30 individus doit être regardé avec méfiance — c'est le seuil du chapitre 8. Ici il tient : 30 observations sur 300, ce n'est pas marginal, et le contenu s'interprète sans effort. **Les retraités ne viennent ni pour le prix ni pour le choix, mais parce que c'est à côté.**

Les cosinus carrés valident les commentaires

Toutes les qualités de représentation dans le plan sont comprises entre 0,76 et 1,00 : les huit points se commentent sans réserve.

Le plus faible est l'indépendant (0,76) : un quart de sa spécificité se trouve sur le troisième axe, qu'on

Les aides à l'interprétation en AFC

Contribution et cosinus carré

$$\text{ctr}_k(i) = \frac{f_{i.} F_k(i)^2}{\lambda_k} \qquad \cos_k^2(i) = \frac{F_k(i)^2}{d_{\chi^2}^2(i, \text{profil moyen})}$$

Le poids $f_{i.}$ remplace $1/n$: c'est la seule différence.

La conséquence de ce changement de poids

En ACP, tous les individus pèsent pareil. En AFC, **une modalité rare pèse peu.**

L'étudiant est le point le plus éloigné de l'origine sur l'axe 1, à $-0,762$ — plus loin que le cadre, à $0,594$.
Il contribue pourtant deux fois moins : $18,8$ contre $37,3\%$, parce qu'il ne pèse que $7,3\%$ de l'effectif.

Le point le plus visible n'est pas le plus déterminant.

Le piège classique de l'AFC

Une modalité à très faible effectif a un profil instable et une distance au centre énorme : elle **s'envole** au bord du graphique, écrasant tout le reste vers l'origine.

Trois remèdes, dans l'ordre : la **regrouper** avec une modalité voisine — l'équivalence distributionnelle garantit que c'est licite si les profils se ressemblent; la passer en **supplémentaire**; ou l'écarter en le disant.

Ce qu'on ne fait pas : commenter un point qui pèse 1 % de l'effectif comme s'il structurait le tableau.

Une ligne ou une colonne projetée

Une ligne supplémentaire se projette par **la même formule barycentrique**, avec son propre profil.

Elle n'entre ni dans le calcul des marges, ni dans celui des axes : sa contribution est nulle par construction.

L'usage le plus utile

Une nouvelle vague d'enquête se projette sur les axes de la précédente. Si le point tombe au même endroit, **la structure du marché n'a pas bougé** ; s'il se déplace, la direction du déplacement dit vers quel motif la clientèle glisse.

C'est ainsi qu'on fait de l'AFC un outil de suivi, et non un cliché isolé.

Ce qu'il faut retenir

Six points

1. Chaque ligne est le **barycentre des colonnes** pondéré par son profil, et réciproquement.
2. Le facteur $1/\sqrt{\lambda_k}$ **dilate** les deux nuages pour les mettre à la même échelle; ici 2,100.
3. En AFC — et là seulement — **la lecture conjointe lignes et colonnes est légitime**.
4. Proximité signifie **sur-représentation par rapport à l'indépendance**, non fréquence élevée.
5. Contribution et cosinus carré sont ceux de l'ACP, le poids $1/n$ devenant la marge $f_{i.}$.
6. Sur nos données : l'axe 1 est la **contrainte budgétaire**, l'axe 2 isole les **retraités** et la proximité.

La suite

L'AFC croise **deux** variables qualitatives. Un questionnaire en compte quinze. Le chapitre 11 étend la méthode à un nombre quelconque de variables : c'est l'analyse des correspondances multiples.