

Analyse des données avancée

Chapitre 11 : L'analyse des correspondances multiples



Jaouad Madkour

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

Le tableau à analyser

Ce qui change par rapport à l'AFC

La lecture

Ce qu'il faut retenir

Le tableau à analyser

Le tableau disjonctif complet

client	fréquence			carte		motif			
	Rare	Mens.	Hebd.	Oui	Non	Prix	Choix	Prox.	Serv.
1	0	0	1	1	0	0	1	0	0
2	1	0	0	0	1	1	0	0	0
3	0	1	0	1	0	0	0	0	1
⋮									
300	0	0	1	1	0	0	0	1	0
somme	1 par ligne			1		1			

Le tableau disjonctif complet : une colonne par modalité, un seul 1 par variable et par ligne.

Chaque ligne totalise donc exactement p , le nombre de variables. Le tableau se comporte alors comme un tableau de contingence — et l'ACM n'est rien d'autre qu'une AFC appliquée à ce tableau.

C'est la même mécanique que les variables indicatrices de l'économétrie, à une différence près : ici on garde *toutes* les modalités, aucune n'est prise comme référence.

Analyse des correspondances multiples

Une AFC appliquée au tableau disjonctif complet de p variables qualitatives.

Aucune théorie nouvelle : la même distance du khi-deux, la même décomposition, les mêmes règles de lecture.

Nos cinq variables

CSP (5 modalités), fréquence (3), carte (2), motif (4), sexe (2).

Total : $K = 16$ colonnes pour $p = 5$ variables et $n = 300$ lignes. Là où l'AFC exigeait dix tableaux, l'ACM n'en demande qu'un.

Le tableau de Burt

	Rare	Mens.	Hebd.	Oui	Non	Prix	Choix	Prox.	Serv.
Rare	112	0	0	26	86	32	28	23	29
Mens.	0	87	0	57	30	21	28	21	17
Hebd.	0	0	101	84	17	31	26	24	20
Oui	26	57	84	167	0	46	47	40	34
Non	86	30	17	0	133	38	35	28	32
Prix	32	21	31	46	38	84	0	0	0
Choix	28	28	26	47	35	0	82	0	0
Prox.	23	21	24	40	28	0	0	68	0
Serv.	29	17	20	34	32	0	0	0	66

Le tableau de Burt : tous les croisements deux à deux, empilés.

Hors diagonale, chaque bloc est un tableau de contingence ordinaire — *fréquence* $\times$ *carte*, *fréquence* $\times$ *motif*. Sur la diagonale, les effectifs de chaque modalité.

Il donne les mêmes axes que le tableau disjonctif, à une transformation près des valeurs propres. On le préfère quand le nombre d'individus est très grand : sa taille ne dépend que des modalités.

Sa lecture directe est illisible — 16×16 ici, davantage en pratique. **C'est précisément pour cela que l'ACM existe.**

Ce qui change par rapport à l'AFC

Une formule qui ne dépend que du comptage

$$I_{\text{tot}} = \frac{K}{p} - 1$$

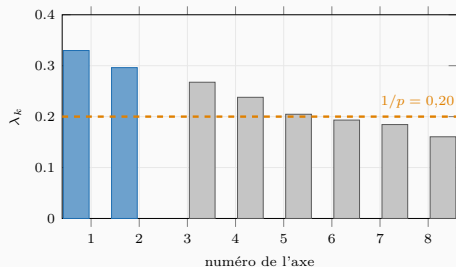
Ici $16/5 - 1 = 2,20$. Cette quantité ne dépend pas des données : elle ne dépend que du nombre de modalités.

La conséquence est sévère

Ajoutez une modalité — découpez l'âge en cinq tranches au lieu de trois — et l'inertie totale augmente, **sans qu'aucune information nouvelle n'entre dans le fichier.**

Les pourcentages d'inertie de l'ACM sont donc **structurellement bas et non comparables** d'une analyse à l'autre. Il faut le savoir avant de lire le premier tableau de sortie.

Les valeurs propres, et comment ne pas s'en effrayer



Ne jugez jamais une ACM sur ses pourcentages.

Le premier axe n'affiche que 15 %. Sur une ACP, ce serait un échec ; ici c'est normal.

Pourquoi : l'inertie totale d'une ACM vaut $K/p - 1$ — ici $16/5 - 1 = 2,20$ — et cette quantité augmente mécaniquement avec le nombre de modalités, sans que l'information augmente.

La règle de Benzécri : ne retenir que les axes dont $\lambda_k > 1/p$. Ici $1/5 = 0,20$, et **cinq axes** passent le seuil. On les interprète par les contributions, jamais par les pourcentages.

Le seuil

Ne retenir que les axes tels que $\lambda_k > 1/p$.

Ici $1/5 = 0,20$, et cinq axes passent le seuil : 0,330, 0,296, 0,268, 0,238, 0,205.

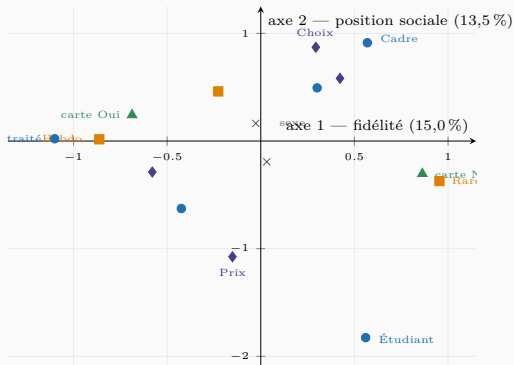
Pourquoi ce seuil-là

Un axe qui ne dépasse pas $1/p$ n'explique pas plus qu'une modalité prise au hasard.

Benzécri proposait en outre une **correction** des valeurs propres retenues, qui rehausse fortement les pourcentages affichés. On la mentionne sans l'appliquer ici : **le principe à retenir est qu'on interprète une ACM par ses contributions, jamais par ses pourcentages.**

La lecture

Le plan des modalités



Lire une ACM, c'est lire des contributions.

Axe 1 : Rare 20,6 %, carte Non 20,0 %, Hebdomadaire 15,2 %, carte Oui 16,0 %. À elles quatre, ces modalités font 72 % de l'axe. **C'est l'axe de la fidélité.**

Axe 2 : Prix 21,8 % et Étudiant 16,5 % contre Choix 14,0 % et Cadre 13,5 %.

Le sexe contribue pour 0,0 % aux deux axes. Il est au centre, et il faut le dire : cette variable n'apporte rien.

Axe 1 : la fidélité

Quatre modalités font **71,8 %** de l'axe : *Rare* (20,6 %), *carte Non* (20,0 %), *carte Oui* (16,0 %), *Hebdomadaire* (15,2 %).

Deux variables sur cinq construisent donc presque tout le premier axe. **C'est l'axe de l'attachement à l'enseigne**, et il recoupe l'axe 1 de l'ACP — vu du côté des comportements plutôt que des opinions.

Axe 2 : la position sociale et le motif

Prix (21,8 %) et *Étudiant* (16,5 %) d'un côté; *Choix* (14,0 %) et *Cadre* (13,5 %) de l'autre.

C'est l'axe 1 de l'AFC du chapitre 10, retrouvé dans une analyse qui ne l'avait pas cherché. Deux méthodes différentes, sur le même fichier, aboutissent au même constat : c'est ce qui donne confiance dans le résultat.

Une variable qui ne dit rien

Le sexe contribue pour zéro pour cent

Femme : $F_1 = -0,027$, $F_2 = +0,164$. Homme : $F_1 = +0,032$, $F_2 = -0,193$. Contributions arrondies à 0,0 % sur les deux axes.

Les deux modalités sont **collées à l'origine**.

Il faut l'écrire noir sur blanc

Une modalité au centre du plan est une modalité dont le profil est **le profil moyen** : elle ne distingue rien.

Le sexe ne structure ni la fréquentation, ni le motif de visite, ni la fidélité. Le chapitre 8 avait déjà montré qu'il ne structurerait pas la satisfaction. Deux méthodes, deux fois le même résultat négatif — et une segmentation par sexe serait, sur ces données, sans fondement.

Ce qu'il faut retenir

Six points

1. L'ACM est **une AFC du tableau disjonctif complet** : aucune notion nouvelle.
2. Le **tableau de Burt** donne les mêmes axes et se préfère quand n est très grand.
3. L'inertie totale vaut $K/p - 1$: elle **ne dépend que du nombre de modalités**.
4. Les pourcentages sont donc **structurellement bas** et non comparables d'une étude à l'autre.
5. **Règle de Benzécri** : retenir les axes tels que $\lambda_k > 1/p$ — ici cinq.
6. On interprète par les **contributions**. Une modalité au centre ne distingue rien : c'est le cas du sexe.

La suite

Nous savons décrire : un plan pour les notes, un plan pour les profils, un plan pour les modalités.

Décrire n'est pas décider. Une direction commerciale ne pilote pas trois cents clients un par un : elle en veut quelques groupes. C'est la quatrième partie.