

Analyse des données avancée

Chapitre 8 : Contributions, cosinus carrés et éléments supplémentaires



Jaouad Madkour

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

Deux questions à ne jamais confondre

La contribution

Le cosinus carré

Les éléments supplémentaires

Ce qu'il faut retenir

Deux questions à ne jamais
confondre

Les deux questions

La **contribution** répond à : *cet élément a-t-il fabriqué l'axe ?*

Le **cosinus carré** répond à : *cet élément est-il bien représenté sur l'axe ?*

Pourquoi elles sont indépendantes

Un individu extrême dans une direction que l'axe ne suit pas : **contribution faible, cosinus carré faible**.

Un individu très excentré exactement dans la direction de l'axe : les deux fortes.

Un individu proche du centre mais parfaitement aligné : **cosinus carré fort, contribution nulle**. Il est bien décrit par l'axe, mais n'y a rien apporté.

Les deux indicateurs se lisent donc **ensemble**, jamais l'un sans l'autre.

La contribution

Définition

Contribution d'un individu à un axe

$$\text{ctr}_k(i) = \frac{p_i F_{ik}^2}{\lambda_k}$$

Avec des poids uniformes : $\text{ctr}_k(i) = F_{ik}^2 / (n\lambda_k)$.

Ses deux propriétés

$$\sum_{i=1}^n \text{ctr}_k(i) = 1 \quad \text{contribution moyenne} = \frac{1}{n}$$

Elle se lit en colonne : la somme des contributions d'un axe vaut 100 %.

Le seuil de lecture

Avec 300 clients, la contribution moyenne vaut $1/300 = 0,33$ %.

L'usage retient les individus dont la contribution dépasse **deux ou trois fois la moyenne** — ici, 1 %. En dessous, l'individu ne dit rien de l'axe.

Les cinq individus les plus contributifs de l'axe 1

Client	F_1	F_2	Contribution
14	-4,99	0,70	1,83 %
168	-4,80	-0,48	1,70 %
293	-4,76	-0,48	1,66 %
166	4,61	-0,84	1,56 %
69	-4,57	-0,46	1,53 %

Application économique

Le plus gros contributeur pèse 1,83 % — cinq fois la moyenne, mais **très loin d'être déterminant**.

Il n'y a donc **aucun individu dominant** : l'axe 1 est porté par l'ensemble du fichier. C'est le résultat qu'on espère, et il faut le vérifier avant tout commentaire.

Si un seul client avait porté 30 % d'un axe, cet axe n'aurait décrit que lui : il aurait fallu le retirer, relancer l'analyse, et le réintroduire comme illustratif.

Le cosinus carré

La qualité de représentation

$$\cos_k^2(i) = \frac{F_{ik}^2}{d^2(i, \mathbf{g})}$$

C'est le carré du cosinus de l'angle entre $\vec{\mathbf{g}}_i$ et l'axe k .

Ses deux propriétés

$$\sum_{k=1}^r \cos_k^2(i) = 1 \qquad \cos_{\text{plan}}^2(i) = \cos_1^2(i) + \cos_2^2(i)$$

Il se lit en ligne, et les qualités s'additionnent d'un axe à l'autre parce que les axes sont orthogonaux.

Intuition

$d^2(i, \mathbf{g})$ est le carré de l'hypoténuse, F_{ik}^2 celui d'un côté. Le rapport est la part de l'éloignement de l'individu que l'axe restitue.

Un cosinus carré de 0,94 : 94 % de ce qui distingue ce client de la moyenne est visible sur le plan. Un cosinus carré de 0,15 : le point qu'on voit sur le graphique **n'est pas ce client**, mais une ombre trompeuse.

note	r axe 1	r axe 2	$\cos^2$ plan	ctr 1	ctr 2
prix	0,629	0,624	0,784	8,7 %	23,6 %
choix	0,771	0,438	0,787	13,1 %	11,7 %
qualité	0,803	0,380	0,789	14,2 %	8,8 %
agencement	0,828	0,242	0,744	15,1 %	3,6 %
propreté	0,827	-0,127	0,701	15,1 %	1,0 %
rapidité	0,742	-0,476	0,778	12,2 %	13,8 %
conseil	0,711	-0,531	0,787	11,1 %	17,1 %
accueil	0,689	-0,580	0,811	10,5 %	20,5 %
total				100 %	100 %

Deux colonnes, deux questions différentes — et on les confond sans cesse.

Le $\cos^2$ répond à : *cette variable est-elle bien représentée sur le plan ?* Il se lit par ligne, et vaut au plus 1.

La **contribution** répond à : *cette variable a-t-elle fabriqué cet axe ?* Elle se lit par colonne, et les colonnes somment à 100 %.

Une variable peut être très bien représentée sans avoir construit l'axe — et une contribution moyenne vaut ici $100/8 = 12,5\%$: au-delà, la variable compte.

Exemple

En **ligne** : la colonne $\cos^2$ du plan. Toutes les valeurs sont entre 0,70 et 0,81 — les huit notes sont **bien représentées**, aucune n'est près du centre du cercle, et toutes les positions relatives se commentent.

En **colonne** : les contributions. Sur l'axe 1, elles vont de 8,7 à 15,1 %, autour d'une moyenne de 12,5 %. **Aucune variable ne fabrique l'axe à elle seule** : c'est bien un effet de taille collectif.

Application économique

Les contributions s'étalent de 1,0 à 23,6 %. Le prix (23,6), l'accueil (20,5) et le conseil (17,1) construisent l'axe; la propreté (1,0) n'y est pour rien.

La propreté est unanime : elle sépare les clients satisfaits des autres, mais ne distingue pas ceux qui viennent pour l'offre de ceux qui viennent pour le service.

Les éléments supplémentaires

Projeter sans participer

Un élément **supplémentaire** — ou illustratif — est projeté sur des axes construits sans lui. Il reçoit des coordonnées et un cosinus carré, mais **jamais de contribution** : sa contribution est nulle par construction.

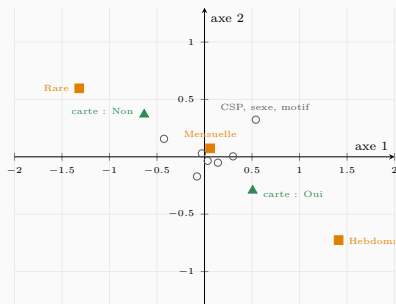
Les trois cas

Variable quantitative : sa coordonnée est sa corrélation avec la composante, et elle se place dans le cercle.

Modalité qualitative : elle se place au **barycentre** des individus qui la portent.

Individu supplémentaire : $\mathbf{F}_{\text{new}} = \mathbf{z}'_{\text{new}} \mathbf{u}_k$, en réduisant avec les moyennes et écarts-types de l'analyse d'origine.

Ce qu'elles révèlent sur nos données



Deux variables parlent, cinq se taisent.

La **fréquences** s'étire le long de l'axe 1, de $-1,32$ à $+1,41$: *le client satisfait revient*. La **carte de fidélité** raconte la même histoire.

La CSP, le sexe et le motif de visite s'agglutinent au centre. Leurs coordonnées ne dépassent pas 0,5.

C'est un résultat, pas un échec. La satisfaction ne dépend pas de qui l'on est. Un service marketing qui segmente par âge et par CSP passe à côté de ce qui structure réellement sa clientèle.

Exemple

Modalité	Effectif	F_1
Rare	112	-1,32
Mensuelle	87	0,06
Hebdomadaire	101	1,41
Femme	162	-0,03
Homme	138	0,03

Application économique

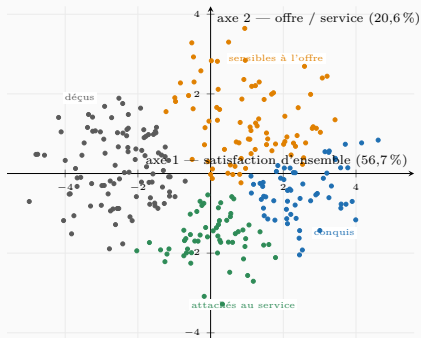
La fréquence s'ordonne parfaitement le long de l'axe 1 — rare, mensuelle, hebdomadaire — alors qu'elle n'a participé à aucun calcul. L'axe de satisfaction **est** un axe de fréquentation. C'est le résultat commercial du fichier.

Le sexe est confondu avec l'origine : $-0,03$ contre $0,03$, sur une échelle où les individus vont de -5 à $+5$. Segmenter par sexe n'aurait, sur ces données, aucun fondement.

Intuition

La deuxième conclusion n'est pas un échec de l'analyse. C'est une **économie** : elle dispense de construire une politique commerciale par sexe, et elle est aussi solide que la première.

Le plan des individus



Le même plan, vu du côté des clients.

Le cercle des corrélations disait ce que *signifient* les axes. Ce plan-ci dit *qui* se trouve où.

À droite : les clients satisfaits de tout. **À gauche** : les mécontents.

En haut : ceux qui louent l'offre et taisent le service. **En bas** : l'inverse.

Les quatre couleurs ne sont pas dans les données : elles viendront de la classification, au chapitre 14. Le nuage, lui, les suggérait déjà.

Ce qu'il faut retenir

Six points

1. **Contribution** : cet élément a-t-il fabriqué l'axe ? Elle se lit **en colonne** et somme à 100 %.
2. **Cosinus carré** : cet élément est-il bien représenté ? Il se lit **en ligne** et somme à 1 sur tous les axes.
3. Les deux sont **indépendants** et se consultent ensemble.
4. Seuil usuel de contribution : deux à trois fois la moyenne $1/n$, soit 1 % ici.
5. Aucun de nos 300 clients ne dépasse 1,9 % : **l'axe 1 n'est porté par personne en particulier.**
6. Les éléments **supplémentaires** se projettent sans participer — d'où la valeur probante de la fréquence, qui s'ordonne toute seule.

La suite

L'ACP est close. Restent les variables **qualitatives**, que ni la moyenne ni la corrélation ne savent traiter : le chapitre 9 leur consacre une métrique propre, celle du khi-deux.