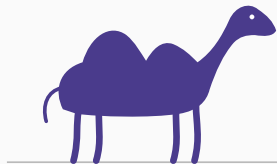


Introduction à l'analyse des données

Chapitre 8 : L'idée de l'analyse en composantes principales



Jaouad Madkour

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

Le problème, et sa solution

Les composantes principales

Le changement de regard

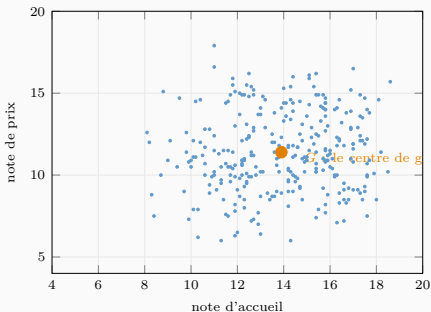
Ce qu'il faut retenir

Le problème, et sa solution

Le programme, rappelé

Trouver la direction de l'espace sur laquelle la **projection du nuage** conserve le maximum d'inertie.

C'est le problème du photographe, énoncé mathématiquement.



Trois objets, et ils suffisent.

Le nuage : 300 points dans un espace à 8 dimensions. On n'en voit ici que deux.

Le centre de gravité G : le client moyen. Centrer les données, c'est le placer à l'origine.

L'inertie : la somme des carrés des distances à G . C'est la *dispersion totale* du nuage — et c'est la quantité que toute l'analyse cherche à conserver.

Intuition

Imaginons un axe passant par le centre du nuage, qu'on fait pivoter.

Pour chaque position, on projette les points et on mesure la dispersion des projections : **grande** quand l'axe suit l'allongement du nuage, **petite** quand il lui est perpendiculaire.

Il existe donc une **position optimale**, et c'est elle qu'on retient.

Les composantes principales

Première composante principale

La **première composante principale** est la direction qui maximise l'inertie projetée.

C'est la direction d'**allongement maximal** du nuage.

Sur le fil rouge

Les cinq clients sont presque alignés – $r = 0,9$. La direction d'allongement est donc la diagonale, celle qui monte de l'individu 1 vers l'individu 5.

Cet axe unique résume à lui seul l'essentiel du tableau : **les clients se distinguent surtout par une satisfaction globale**, qui fait varier ensemble l'accueil et la fidélité.

Les axes suivants

Les composantes suivantes

La **deuxième** composante est la direction qui maximise l'inertie **restante**, sous une contrainte : elle doit être **orthogonale** à la première.

Puis la troisième, orthogonale aux deux premières, et ainsi de suite jusqu'à p axes.

Pourquoi l'orthogonalité

Sans cette contrainte, le deuxième axe se placerait presque au même endroit que le premier : il redirait la même chose.

L'orthogonalité garantit que **chaque nouvel axe apporte une information que les précédents n'avaient pas**. Les composantes sont ainsi non corrélées entre elles, et l'inertie qu'elles portent s'additionne sans double compte.

Trois propriétés à retenir

1. Les axes sont **hiérarchisés** : le premier porte le plus d'inertie, puis le deuxième, et ainsi de suite.
2. Ils sont **orthogonaux** deux à deux.
3. Les inerties portées **s'additionnent** et redonnent l'inertie totale.

Le changement de regard

Ce qu'est une composante

Une composante principale est une **combinaison linéaire** des variables de départ :

$$C_1 = a_1 z_1 + a_2 z_2 + \dots + a_p z_p$$

C'est une variable **nouvelle**, qu'on n'a pas mesurée mais construite.

Ce que cela donne en gestion

Sur un fichier de satisfaction, le premier axe combine souvent positivement **tous** les critères : c'est un **niveau général de satisfaction**, que personne n'a mesuré directement.

Le deuxième oppose fréquemment deux familles de critères — par exemple le service contre le prix. Il mesure alors un **arbitrage** entre deux dimensions du jugement.

L'analyse fabrique ainsi des indicateurs synthétiques, et c'est souvent le principal résultat pratique de l'exercice.

Intuition

Ces variables construites n'existaient pas dans le tableau. Rien ne garantit qu'elles aient un sens.

Les interpréter est un travail à part entière : il faut regarder quelles variables les composent — c'est le chapitre 11 — et connaître le domaine. Une composante mal nommée est une erreur durable.

Quand l'ACP est utile, et quand elle ne l'est pas

La condition

L'ACP n'apporte quelque chose que si les variables sont **corrélées entre elles**.

Si toutes les corrélations sont nulles, chaque variable porte une information propre : aucune réduction n'est possible, et tous les axes portent la même inertie.

Le contrôle préalable

D'où l'importance de la **matrice des corrélations** du chapitre 5 : elle se regarde avant de lancer l'analyse.

Une matrice ne contenant que des corrélations proches de zéro annonce que l'ACP ne servira à rien.

Une matrice pleine de valeurs fortes annonce l'inverse — les variables se répètent, et l'on pourra les résumer.

Ce qu'il faut retenir

Cinq points

1. La **première composante** est la direction qui maximise l'inertie projetée.
2. Les suivantes maximisent l'inertie **restante**, sous contrainte d'**orthogonalité**.
3. Les axes sont **hiérarchisés**, orthogonaux, et leurs inerties **s'additionnent**.
4. Une composante est une **combinaison linéaire** des variables — une variable construite, à interpréter.
5. L'ACP n'est utile que si les variables sont **corrélées** : la matrice des corrélations le dit d'avance.

La suite

Reste à savoir **comment** ces axes se calculent. Le chapitre 9 donne la réponse — et elle tient dans un objet du cours d'algèbre linéaire.