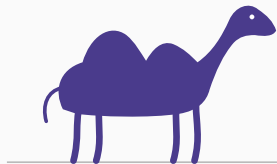


Introduction à l'analyse des données

Chapitre 10 : Combien d'axes retenir



Jaouad Madkour

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

La question

Les trois critères

Le plan factoriel

Ce qu'il faut retenir

La question

Les deux extrêmes sont mauvais

Retenir **tous** les axes ne perd aucune information — et ne réduit rien. On revient au tableau de départ, avec son problème de dimension intact.

Retenir **un seul** axe donne la représentation la plus simple, au prix d'une perte qui peut être considérable.

Il **n'existe pas de réponse unique**, et aucun critère ne tranche à lui seul. On en utilise donc trois, et l'on regarde s'ils concordent.

Les trois critères

La règle

Sur données réduites, retenir les axes dont la valeur propre dépasse 1 :

$$\lambda_k > 1$$

Sa justification

Chaque variable réduite porte exactement une unité d'inertie. Un axe qui en porte moins **résume moins bien qu'une variable seule** — il ne mérite donc pas d'être conservé.

Le seuil n'est pas arbitraire : il découle directement du fait que $I = p$, établi au chapitre 7.

Ses limites

La règle est mécanique et souvent trop généreuse quand les variables sont nombreuses. Un axe à 1,02 et un axe à 0,98 ne diffèrent en rien de substantiel, et pourtant la règle les sépare.

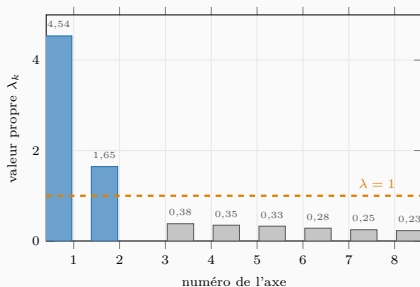
C'est un point de départ, pas une décision.

Le critère du coude

L'éboulis des valeurs propres

On trace les valeurs propres par ordre décroissant, et l'on cherche le **décrochement** — le point où la courbe cesse de chuter et s'aplatit.

On retient les axes situés **avant** ce coude.



Les trois critères disent la même chose.

Le coude : la falaise entre l'axe 2 et l'axe 3 — de 1,65 à 0,38. On coupe au pied de la falaise.

Kaiser : garder les axes de valeur propre > 1 , c'est-à-dire ceux qui résument mieux qu'une variable seule. Ici : deux.

Le cumul : $56,7 + 20,6 = 77,3\%$ de l'information conservée sur un simple plan.

Quand les trois concordent, la décision est facile. Quand ils divergent, c'est l'interprétabilité qui tranche.

Ce que le coude traduit

Avant le coude, chaque axe apporte nettement moins que le précédent : les dimensions sont

La règle

Retenir assez d'axes pour atteindre un seuil d'inertie cumulée — souvent 70 ou 80 %.

Le seuil dépend du nombre de variables

Avec **trois** variables, atteindre 80 % est facile et ne prouve rien.

Avec **trente**, y parvenir sur trois axes est une réduction remarquable — de trente dimensions à trois, en ne perdant qu'un cinquième.

Un **pourcentage** ne se juge donc jamais dans l'**absolu**, mais rapporté au p de départ.

Un cas typique, sur huit variables

Axe	λ	Part	Cumul
1	3,2	40 %	40 %
2	1,8	22,5 %	62,5 %
3	1,1	13,8 %	76,3 %
4	0,7	8,8 %	85,1 %
5	0,5	6,3 %	91,4 %

- **Kaiser** retient trois axes (3,2; 1,8; 1,1 dépassent 1).
- Le **coude** se situe après le troisième : la chute 1,1 $\rightarrow$ 0,7 est nette, puis la décroissance ralentit.
- Le **cumul** atteint 76,3 % sur trois axes.

Les trois critères concordent : on retient trois axes.

Intuition

C'est le cas fréquent. Il faut alors ajouter un quatrième critère, qui n'est pas numérique : **l'axe est-il interprétable ?**

Un axe à 1,05 sur lequel aucune variable ne se distingue n'apporte rien. Un axe à 0,95 qui oppose clairement deux familles de critères mérite d'être gardé.

C'est un jugement, et il doit être motivé — pas dissimulé derrière un seuil.

Le plan factoriel

Deux axes, une feuille

Le plan principal

En pratique, on représente presque toujours le **plan** formé des deux premiers axes.

Son inertie est $(\lambda_1 + \lambda_2)/p$.

Sur le cas à huit variables

Le premier plan porte $40 + 22,5 = 62,5\%$ de l'inertie.

Si un troisième axe est retenu, on trace un second plan — axes 1 et 3, ou 2 et 3 — plutôt que d'imaginer une figure en trois dimensions.

La règle de lecture la plus importante

Toute lecture d'un plan factoriel doit être accompagnée de son pourcentage d'inertie.

Un plan à 62 % se commente avec prudence : plus d'un tiers de l'information n'y figure pas. Deux individus qui y paraissent proches peuvent être éloignés dans les dimensions non représentées.

Le chapitre 12 donnera l'outil qui permet de vérifier cela individu par individu.

Ce qu'il faut retenir

Cinq points

1. Le nombre d'axes est un **arbitrage** entre simplicité et fidélité.
2. **Kaiser** : garder les axes tels que $\lambda > 1$, puisque chaque variable pèse 1.
3. Le **coude** de l'éboulis sépare la structure du bruit — critère visuel.
4. Le **cumul** se juge relativement au nombre de variables de départ.
5. Un quatrième critère décide en cas de désaccord : **l'axe est-il interprétable ?**

La suite

Les axes sont choisis. Il faut maintenant les **interpréter** — et cela commence par le côté des variables, avec le cercle des corrélations.