

Analyse des données avancée

Chapitre 7 : Combien d'axes retenir



Jaouad Madkour

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

Une question sans réponse unique

Les quatre critères

Prolonger au-delà des critères

Ce qu'il faut retenir

Une question sans réponse
unique

L'arbitrage

Garder tous les axes : aucune perte, aucun gain — on n'a rien résumé.

Garder un seul axe : un résumé maximal, une perte peut-être inacceptable.

Aucun théorème ne tranche. Le choix relève du jugement, éclairé par plusieurs critères convergents.

La règle qui vaut plus que tous les critères

Un axe se garde s'il **s'interprète**. Un axe dont personne ne sait quoi dire ne sert à rien, quel que soit son pourcentage d'inertie.

Inversement, un troisième axe à 5 % mais qui isole nettement un segment de clientèle mérite d'être gardé. Les critères qui suivent orientent la décision ; ils ne la remplacent pas.

Les quatre critères

Le premier réflexe

$$\text{taux cumulé} = \frac{\lambda_1 + \dots + \lambda_k}{I}$$

On retient assez d'axes pour dépasser un seuil : 70, 80 ou 90 % selon l'usage.

Pourquoi ce critère est le plus mal employé

Le seuil « raisonnable » dépend entièrement de p . Avec 3 variables, deux axes donnent facilement 90 % et cela ne signifie rien. Avec 50 variables, 60 % sur deux axes est un résultat remarquable.

Un pourcentage ne se juge jamais dans l'absolu, mais toujours contre ce qu'un tableau sans structure aurait donné — ce que le critère suivant chiffre.

En ACP normée seulement

retenir les axes tels que $\lambda_k > 1$

L'argument, qui est bon

En ACP normée, chaque variable de départ porte une inertie de 1, et le total vaut p . La valeur propre moyenne vaut donc exactement 1.

Un axe de valeur propre inférieure à 1 **résume moins d'information qu'une variable isolée n'en apporte**. Le remplacer par une combinaison de huit variables serait un mauvais échange.

Sur nos données

$\lambda_1 = 4,535$ et $\lambda_2 = 1,645$ dépassent 1. $\lambda_3 = 0,382$ ne l'atteint pas — et de loin.

Deux axes.

Un seuil corrigé

$$\lambda_k > 1 + 2\sqrt{\frac{p-1}{n-1}}$$

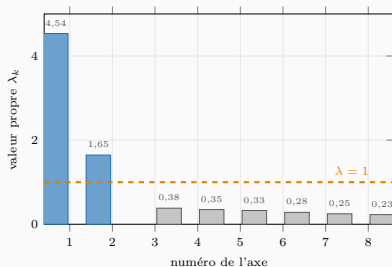
Le critère de Kaiser est connu pour être **trop généreux** quand n est petit devant p .

Sur nos données

$$1 + 2\sqrt{\frac{7}{299}} = 1 + 2 \times 0,153 = 1,306$$

$\lambda_2 = 1,645$ passe encore. **Deux axes**, de nouveau.

L'éboulis et le coude de Cattell



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.

Définition

On repère la cassure dans le graphique des valeurs propres, et l'on retient les axes **avant** elle.

Exemple

Passage	Chute
$\lambda_1 \rightarrow \lambda_2$	2,890
$\lambda_2 \rightarrow \lambda_3$	1,263
$\lambda_3 \rightarrow \lambda_4$	0,033
$\lambda_4 \rightarrow \lambda_5$	0,022
$\lambda_5 \rightarrow \lambda_6$	0,044

La chute s'effondre d'un facteur 40 entre la deuxième et la troisième. **Le coude est net.**

Intuition

Les six dernières valeurs propres vont de 0,382 à 0,229 : elles sont **presque égales**.

C'est la signature d'un nuage sans direction privilégiée — donc du bruit. Un éboulis plat ne se coupe pas au milieu : il se laisse entier de côté.

Prolonger au-delà des critères

Le verdict sur nos données

Critère	Décision
Cumulé à 77,3 %	deux axes suffisent
Kaiser	deux axes
Karlis, Saporta et Spinaki	deux axes
Coude de Cattell	deux axes

Une convergence rassurante, et rare

Quatre critères indépendants donnent la même réponse. C'est le cas favorable, et il n'est pas le plus courant.

Quand ils divergent — Kaiser en donne quatre, le coude en donne deux — **on tranche par l'interprétation**, puis on refait l'analyse avec l'autre nombre pour voir si les conclusions bougent.

Trois fautes

Choisir le nombre d'axes après avoir vu les résultats qui plaisent. C'est de la construction de conclusion, non de l'analyse.

Garder un axe non interprétable parce qu'il dépasse un seuil. Un axe sans nom ne se commente pas.

Retenir un axe porté par un seul individu. Le chapitre 8 donne l'outil qui le détecte : la contribution.

La question n'est pas la même

Pour **représenter**, deux axes s'imposent : on ne dessine pas en dimension trois.

Pour **classer** — au chapitre 12 et après — rien n'oblige à visualiser. L'usage est d'aller jusqu'à 80 ou 90 % d'inertie, quitte à garder cinq ou six axes qu'on ne dessinera jamais.

Ce que la troncature apporte à la classification

Classer sur les axes retenus plutôt que sur les variables brutes n'est pas seulement plus rapide : **c'est plus juste.**

Les derniers axes portent le bruit. Les écarter avant de classer, c'est empêcher le bruit de créer des classes.

Ce qu'il faut retenir

Six points

1. **Aucun théorème** ne dit combien d'axes garder : c'est un jugement.
2. Le **pourcentage cumulé** ne se juge pas dans l'absolu, mais contre le nombre de variables.
3. **Kaiser** : en ACP normée, garder les axes tels que $\lambda_k > 1$, la moyenne des valeurs propres valant 1.
4. Le seuil corrigé $1 + 2\sqrt{(p-1)/(n-1)}$ vaut ici 1,306, et ne change pas la conclusion.
5. Le **coude de l'éboulis** est net entre le deuxième et le troisième axe.
6. Un axe se garde s'il **s'interprète** — ce critère prime sur les quatre autres.

La suite

Deux axes sont retenus. Le chapitre 8 donne les trois outils sans lesquels on ne peut pas les commenter : la **contribution**, le **cosinus carré** et les **éléments supplémentaires**.