

Analyse des données avancée

Chapitre 3 : L'inertie, formalisée



Jaouad Madkour

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

La définition générale

L'inertie sur le petit tableau

Projeter, c'est décomposer

Ce qu'il faut retenir

La définition générale

L'inertie totale

$$I = \sum_{i=1}^n p_i d_M^2(\mathbf{x}_i, \mathbf{g}) = \sum_{i=1}^n p_i (\mathbf{x}_i - \mathbf{g})' \mathbf{M} (\mathbf{x}_i - \mathbf{g})$$

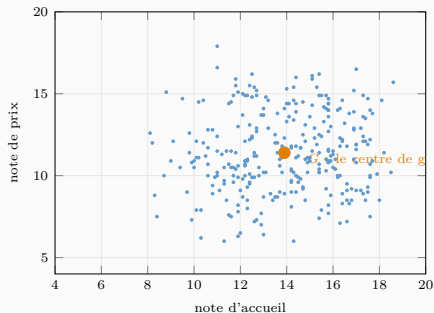
C'est la **moyenne pondérée** des carrés des distances au centre de gravité.

Le mot vient de la mécanique

En physique, le moment d'inertie mesure la résistance d'un solide à la rotation : il est grand quand la masse est loin de l'axe.

Ici, l'inertie est grande quand les individus sont **loin de la moyenne**, c'est-à-dire quand ils se ressemblent peu. **Un nuage d'inertie nulle est un nuage où tout le monde est identique** — et il n'y a rien à analyser.

Le nuage et son centre

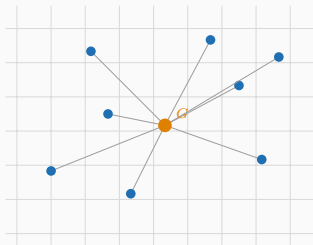


Trois objets, et ils suffisent.

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.



L'inertie totale

$$I_{\text{tot}} = \frac{1}{n} \sum_{i=1}^n d^2(i, G)$$

Elle mesure exactement ce que mesurait la variance — et pour une seule variable, elle *est* la variance.

Sur données centrées-réduites, chaque variable apporte une inertie de 1. Avec p variables, l'inertie totale vaut donc p : ici 8.

C'est la monnaie de tout le cours. « Cet axe explique 56,7 % » signifiera toujours : il en conserve 56,7 %.

La forme matricielle

$$I = \text{trace}(\mathbf{VM}), \quad \mathbf{V} = \mathbf{X}'\mathbf{D}\mathbf{X}$$

où $\mathbf{X}$ est centré et $\mathbf{V}$ est la matrice de variance-covariance.

Pourquoi la démonstration tient en deux lignes

Chaque terme $\mathbf{x}_i'\mathbf{M}\mathbf{x}_i$ est un **scalaire**, donc égal à sa propre trace. Or $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$:

$$\mathbf{x}_i'\mathbf{M}\mathbf{x}_i = \text{trace}(\mathbf{x}_i'\mathbf{M}\mathbf{x}_i) = \text{trace}(\mathbf{M}\mathbf{x}_i\mathbf{x}_i')$$

En sommant avec les poids : $I = \text{trace}(\mathbf{M} \sum_i p_i \mathbf{x}_i \mathbf{x}_i') = \text{trace}(\mathbf{MV}) = \text{trace}(\mathbf{VM})$.

Ce que cette formule apporte

Trois lectures, un seul nombre

En ACP **non normée** avec $\mathbf{M} = \mathbf{I}$:

$$I = \text{trace}(\mathbf{V}) = \sum_{j=1}^p s_j^2$$

En ACP **normée** avec $\mathbf{M} = \mathbf{D}_{1/s^2}$:

$$I = \text{trace}(\mathbf{R}) = p$$

Sur nos données

Huit variables réduites, donc $I = 8$ **exactement**. On le vérifie sur le tableau : la somme des huit valeurs propres vaut

$$4,535 + 1,645 + 0,382 + 0,349 + 0,327 + 0,283 + 0,249 + 0,229 = 8,000$$

L'inertie sur le petit tableau

Quatre individus, deux variables

$$\mathbf{X} = \begin{pmatrix} -2 & -2 \\ -1 & 1 \\ 1 & -1 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$$

Avec $p_i = 1/4$:

$$\mathbf{V} = \frac{1}{4}\mathbf{X}'\mathbf{X} = \begin{pmatrix} 2,5 & 1,5 \\ 1,5 & 2,5 \end{pmatrix}$$

L'inertie, par les deux chemins

Par la trace : $I = 2,5 + 2,5 = 5$.

Par les distances : $\frac{1}{4}[(4+4) + (1+1) + (1+1) + (4+4)] = \frac{20}{4} = 5$.

Les deux coïncident, comme la démonstration l'annonçait.

Intuition

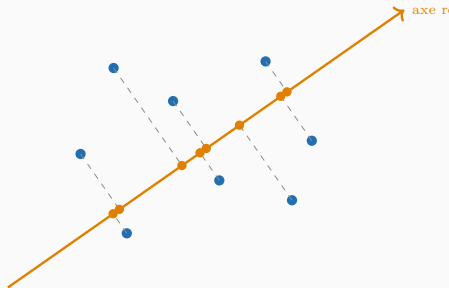
Les deux variances valent 2,5, donc $\mathbf{M} = \text{diag}(0,4 ; 0,4)$ et

$$I = \text{trace}(\mathbf{VM}) = 1 + 1 = 2 = p$$

Le même nuage porte une inertie de 5 ou de 2 **selon la métrique**. Le nombre n'a donc de sens que rapporté au total : c'est pourquoi on ne lit jamais une valeur propre, mais un **pourcentage d'inertie**.

Projeter, c'est décomposer

Le théorème de Pythagore, appliqué au nuage



Projeter, c'est écraser.

Chaque point bleu tombe sur l'axe. Ce qui subsiste : sa position *le long* de l'axe. Ce qui disparaît : sa distance à l'axe — les traits pointillés.

Théorème de Pythagore, appliqué au nuage entier :

$$I_{\text{tot}} = \underbrace{I_{\text{projetée}}}_{\text{conservée}} + \underbrace{I_{\text{résiduelle}}}_{\text{perdue}}$$

Le meilleur axe est donc celui qui maximise l'inertie projetée — ce qui revient exactement à minimiser la somme des carrés des pointillés.

L'égalité qui fonde toute la suite

La décomposition

Pour un sous-espace E et son orthogonal :

$$\underbrace{I}_{\text{totale}} = \underbrace{I_E}_{\text{conservée}} + \underbrace{I_{E^\perp}}_{\text{perdue}}$$

La conséquence, décisive

I ne dépend pas du sous-espace choisi : c'est une propriété du nuage.

Donc **maximiser l'inertie projetée revient exactement à minimiser l'inertie perdue**. Chercher le meilleur résumé ou chercher la plus petite déformation sont le même problème — et c'est pourquoi l'analyse des données n'a qu'un seul critère.

Le chameau, en langage d'inertie

Le photographe qui cherche le meilleur angle **maximise l'inertie projetée**. La silhouette de profil conserve la longueur du corps, celle de face la perd tout entière.

Ce que la Licence disait en images, la formule le dit en un signe d'égalité.

Le problème posé proprement

Trouver l'axe $\mathbf{u}$, normé au sens de $\mathbf{M}$, qui maximise l'inertie projetée :

$$\max_{\mathbf{u}} \mathbf{u}'\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{u} \quad \text{sous} \quad \mathbf{u}'\mathbf{M}\mathbf{u} = 1$$

Ce qu'il faut avant de le résoudre

Un problème de maximisation sous contrainte quadratique se résout par les multiplicateurs de Lagrange, et sa solution s'écrit avec des **vecteurs propres**.

Le chapitre 4 rappelle donc la diagonalisation et introduit la décomposition en valeurs singulières — l'outil qui, seul, explique pourquoi individus et variables donnent les mêmes valeurs propres.

Ce qu'il faut retenir

Cinq points

1. L'**inertie** est la moyenne pondérée des carrés des distances au centre : elle mesure la dispersion du nuage.
2. Elle s'écrit $I = \text{trace}(\mathbf{VM})$, et vaut p en ACP normée.
3. Sur nos données : $I = 8$, et la somme des valeurs propres le confirme.
4. Projeter décompose l'inertie en **conservée** et **perdue**, dont la somme est constante.
5. **Maximiser ce qu'on garde, c'est minimiser ce qu'on perd** : les deux formulations du problème n'en font qu'une.

La suite

Le problème est posé. Le chapitre 4 fournit l'outil qui le résout — et le résout **une fois pour toutes les méthodes du cours**.