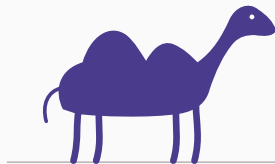


Introduction à l'analyse des données

Chapitre 6 : Le tableau de données et le nuage



Jaouad Madkour

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

Le tableau

Les deux nuages

Le centre de gravité

Ce que le nuage permet de dire

Ce qu'il faut retenir

Le tableau

client	8 notes de satisfaction				signalétique	
	prix	choix	...	accueil	CSP	fréquence
1	13,2	15,4	...	16,1	Cadre	Hebdo.
2	9,8	11,0	...	10,4	Employé	Rare
⋮	⋮	⋮		⋮	⋮	⋮
300	12,5	13,9	...	14,8	Retraité	Mensuelle

variables **actives** : elles construisent l'espace variables **illustratives** : on les projette ensuite

Le choix des variables actives est la seule décision vraiment engageante de toute l'analyse.

Elles fixent la géométrie du nuage — donc les axes, donc l'interprétation. Les illustratives n'y participent pas : on les superpose après coup pour *éclairer* les axes, jamais pour les construire.

Mettre le chiffre d'affaires en actif quand on étudie la satisfaction, c'est laisser la taille des clients écraser leur opinion.

Le tableau individus-variables

n lignes — les **individus** — et p colonnes — les **variables** : la case (i, j) porte la valeur de j pour i . C'est le format d'entrée de toute l'analyse des données.

Application économique

Point	Question
Homogénéité	les lignes sont-elles comparables entre elles?
Valeurs manquantes	combien, et pourquoi manquent-elles?
Valeurs aberrantes	erreurs de saisie, ou individus réellement atypiques?
Doublons	le même client compte-t-il deux fois?

Un individu aberrant peut **capter à lui seul un axe entier** — c'est la sensibilité du chapitre 5, aggravée.
Ce contrôle n'est pas une formalité.

Les deux nuages

Un tableau, deux lectures

Le nuage des individus

Chaque **ligne** devient un **point** dans un espace à p dimensions : ses coordonnées sont ses p valeurs.

Cinq clients notés sur deux critères donnent cinq points dans un plan.

Le nuage des variables

Chaque **colonne** devient un **point** dans un espace à n dimensions.

Deux variables observées sur cinq clients donnent deux points dans un espace à cinq dimensions.



Une seule ACP, deux lectures — c'est ce qu'on appelle la dualité.

On ne fait pas deux calculs : la décomposition de la matrice des corrélations donne *en même temps* les axes des individus et ceux des variables. Le pourcentage d'inertie est donc le même des deux côtés.

D'où la méthode de lecture, invariable : on interprète d'abord le *cercle des corrélations* — il donne le *sens* des axes — puis le *plan des individus* — il dit *qui* se trouve où. Jamais l'inverse.

Intuition

Ce sont deux façons de lire **le même tableau**, et elles répondent à deux questions distinctes :

- le nuage des **individus** demande : quels clients se ressemblent ?
- le nuage des **variables** demande : quels critères vont ensemble ?

Elles se répondront simultanément : c'est la **dualité**, qui donnera le cercle des corrélations au chapitre 11 et le plan des individus au chapitre 12.

Cinq points dans le plan

Client	Accueil	Fidélité
1	8	4
2	10	12
3	12	8
4	14	16
5	16	20

Placés dans un repère, ces cinq points s'alignent presque : c'est la traduction visuelle du $r = 0,9$ du chapitre 5.

Avec deux variables, on voit tout. Le problème du chapitre 1 ne se pose pas encore — il commencera à trois.

Le centre de gravité

Le point moyen

Centre de gravité

Le **centre de gravité** G du nuage est le point dont les coordonnées sont les **moyennes** de chaque variable :

$$G = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$$

Sur le fil rouge

$$G = (12; 12)$$

Les deux moyennes valant 12, le centre du nuage est le point (12 ; 12).

Ce que centrer signifie géométriquement

Centrer les variables, comme au chapitre 3, revient à **déplacer l'origine du repère en G** .

Le nuage ne change ni de forme, ni de taille, ni d'orientation : seul le point d'où on le regarde a bougé.

C'est une préparation nécessaire, car les axes qu'on cherchera **passeront tous par le centre du nuage**.

Un axe qui n'y passerait pas ne pourrait pas résumer l'ensemble.

Après centrage et réduction

Ce que le nuage permet de dire

La distance entre deux individus

Sur données réduites, la distance entre deux individus i et i' se calcule par le théorème de Pythagore, étendu à p dimensions :

$$d(i, i') = \sqrt{\sum_{j=1}^p (z_{ij} - z_{i'j})^2}$$

Ce que la géométrie apporte

Deux clients **proches** dans le nuage ont des valeurs semblables sur l'ensemble des critères. Deux clients **éloignés** diffèrent.

La question « quels clients se ressemblent ? » devient donc une question de **distance** — et ce basculement vers la géométrie rend la suite possible.

Application économique

Sans réduction, la distance serait dominée par la variable de plus grande dispersion : deux clients seraient jugés éloignés parce qu'ils diffèrent en montant dépensé, quelle que soit leur satisfaction.

Réduire, c'est décider que chaque critère compte autant dans la ressemblance. C'est un choix, et il faut savoir qu'on le fait.

Ce qu'il faut retenir

Cinq points

1. Le tableau **individus-variables** est le format d'entrée : n lignes, p colonnes.
2. Il se lit de **deux façons** : nuage des individus dans $\mathbb{R}^p$, nuage des variables dans $\mathbb{R}^n$.
3. Le **centre de gravité** a pour coordonnées les moyennes — ici (12 ; 12).
4. **Centrer** revient à placer l'origine en G ; la forme du nuage ne change pas.
5. La **ressemblance** entre individus devient une **distance**, ce qui rend la géométrie utilisable.

La suite

Le nuage est en place. Reste à mesurer l'information qu'il contient — sans quoi on ne pourra pas dire qu'une projection en conserve beaucoup ou peu.

C'est l'inertie, et c'est le chapitre 7.