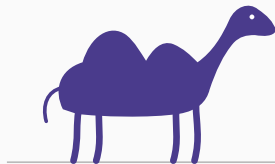


Introduction à l'analyse des données

Chapitre 3 : Rappels sur une variable : centrer et réduire



Jaouad Madkour

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

Le fil rouge du cours

La moyenne et la variance

Centrer et réduire

Ce qu'il faut retenir

Le fil rouge du cours

Cinq clients, deux critères

Le tableau qui portera tout le cours

Cinq clients ont noté l'enseigne sur deux critères, de 0 à 20 :

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

Pourquoi cinq lignes

Un fichier réel en compte des milliers. Cinq permettent de **tout vérifier à la main**, et c'est le seul moyen de comprendre ce que les logiciels calculent.

Ce tableau reviendra à chaque chapitre jusqu'à l'analyse en composantes principales.

La moyenne et la variance

La moyenne

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Sur nos données

$$\bar{x} = \frac{8 + 10 + 12 + 14 + 16}{5} = \frac{60}{5} = 12 \qquad \bar{y} = \frac{4 + 12 + 8 + 16 + 20}{5} = \frac{60}{5} = 12$$

Les deux variables ont la **même moyenne**. Elles ne se ressemblent pourtant pas du tout, et c'est la variance qui va le dire.

Définition

$$V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

C'est la **moyenne des carrés des écarts** à la moyenne. L'**écart-type** $s_x = \sqrt{V(x)}$ la ramène à l'unité de la variable.

Le calcul, écart par écart

Les cinq écarts, et leurs carrés

Client	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	-4	16	-8	64
2	-2	4	0	0
3	0	0	-4	16
4	2	4	4	16
5	4	16	8	64
Somme	0	40	0	160

$$V(x) = \frac{40}{5} = 8 \quad V(y) = \frac{160}{5} = 32$$

$$s_x = \sqrt{8} \approx 2,83 \quad s_y = \sqrt{32} \approx 5,66$$

Centrer et réduire

L'opération qui commande tout le cours

La variable centrée réduite

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

Centrer, c'est retrancher la moyenne. **Réduire**, c'est diviser par l'écart-type.

Sur l'accueil

Client	x_i	$x_i - 12$	$z_i = \frac{x_i - 12}{2,83}$
1	8	-4	-1,414
2	10	-2	-0,707
3	12	0	0
4	14	2	0,707
5	16	4	1,414

Moyenne nulle, variance unité

$$\bar{z} = 0 \qquad V(z) = 1$$

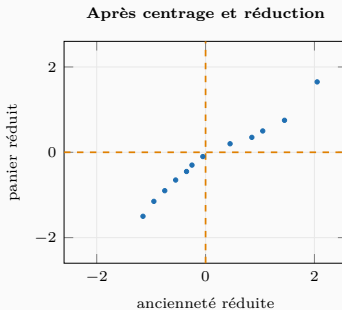
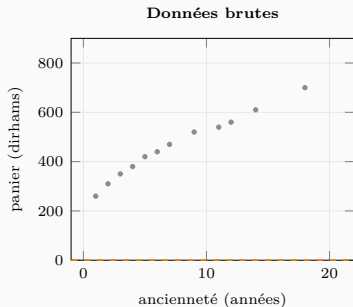
Toute variable centrée réduite a une moyenne nulle et une variance de un, quelles que soient ses valeurs de départ.

Pourquoi c'est indispensable ici

Le problème des unités

Un fichier client mêle des variables sans commune mesure : un montant en dirhams qui varie de 200 à 50 000, une note de satisfaction entre 1 et 5, une ancienneté en années.

La variance du montant se compte en millions; celle de la note, en dixièmes. Sans précaution, **le montant écraserait tout** : compter en milliers de dirhams plutôt qu'en dirhams changerait les conclusions. C'est inacceptable.



Ce qu'il faut retenir

Cinq points

1. Le fil rouge : **cinq clients**, accueil et fidélité, moyennes toutes deux égales à 12.
2. $V(x) = 8$ et $V(y) = 32$: même moyenne, dispersions très différentes.
3. La somme des écarts à la moyenne vaut **toujours zéro** — c'est un contrôle.
4. La variable **centrée réduite** $z = (x - \bar{x})/s_x$ a une moyenne de 0 et une variance de 1.
5. Réduire donne à chaque variable **le même poids** et supprime l'effet des unités.

La suite

Une variable est décrite. Il faut maintenant en relier **deux** — et le chapitre 4 montre que l'outil pour le faire n'est pas nouveau : c'est la variance, généralisée.