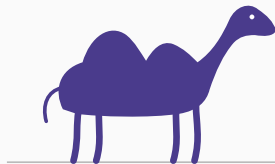


Introduction à l'analyse des données

Chapitre 4 : La covariance, généralisation de la variance



Jaouad Madkour

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

D'une variable à deux

Le calcul

La limite de la covariance

La matrice des variances-covariances

Ce qu'il faut retenir

D'une variable à deux

La variance, réécrite

Le point de départ

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

Écrite ainsi, la variance est la **moyenne du produit de l'écart de x par lui-même**.

La question qui vient naturellement

Si l'on peut multiplier l'écart de x par l'écart de x , rien n'interdit de le multiplier par l'écart de y .

C'est exactement ce que fait la covariance. Elle n'est pas une notion nouvelle : c'est la variance, avec une seconde variable à la place de la première.

La covariance

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})$$

La conséquence immédiate

Le calcul

Produit des écarts, client par client

Client	$x_i - \bar{x}$	$y_i - \bar{y}$	Produit
1	-4	-8	+32
2	-2	0	0
3	0	-4	0
4	2	4	+8
5	4	8	+32
Somme			72

$$\text{cov}(x, y) = \frac{72}{5} = 14,4$$

Intuition

Les clients 1 et 5 sont du même côté de la moyenne sur les deux critères — deux écarts de même signe, donc un produit **positif**. Un client au-dessus sur l'un et en dessous sur l'autre donnerait un produit **négatif**.

La **covariance additionne ces accords et ces désaccords**. Elle est positive lorsque les deux variables varient dans le même sens.

Ce que le signe signifie

Les trois cas

Signe	Lecture
$\text{cov} > 0$	les deux variables varient dans le même sens
$\text{cov} < 0$	elles varient en sens contraire
$\text{cov} = 0$	aucune liaison linéaire

Sur nos clients

$\text{cov}(x, y) = 14,4 > 0$: les clients qui jugent bien l'accueil sont aussi les plus fidèles.

C'est une information de gestion immédiate — et c'est encore une **association**, non une causalité. Rien ici ne dit que soigner l'accueil produira de la fidélité.

La limite de la covariance

Un nombre qu'on ne sait pas interpréter

Quatorze virgule quatre, est-ce beaucoup ?

Impossible à dire. La covariance dépend des **unités** : elle s'exprime ici en « points d'accueil $\times$ points de fidélité », ce qui ne veut rien dire.

Si les notes étaient sur 100 au lieu de 20, la covariance serait multipliée par 25 — sans qu'aucune liaison ait changé.

Aucune borne

La covariance n'a ni maximum ni minimum. Elle peut valoir 0,003 ou 4 500 000 selon les grandeurs manipulées.

On ne peut donc **ni la comparer d'un couple de variables à l'autre**, ni juger de la force d'une liaison à partir d'elle.

Le problème est celui du chapitre 3

C'est exactement la difficulté des unités déjà rencontrée. Et le remède sera le même : **réduire**.

Diviser par les écarts-types va supprimer les unités et borner le résultat. C'est le chapitre 5, et cela donnera la corrélation.

La matrice des variances-covariances

Ranger tous les calculs

La matrice

Avec p variables, on range les variances sur la **diagonale** et les covariances ailleurs :

$$\Sigma = \begin{pmatrix} V(x_1) & \text{cov}(x_1, x_2) & \cdots \\ \text{cov}(x_2, x_1) & V(x_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Sur le fil rouge

$$\Sigma = \begin{pmatrix} 8 & 14,4 \\ 14,4 & 32 \end{pmatrix}$$

La diagonale porte les variances du chapitre 3; les deux autres cases portent la covariance.

Deux propriétés

1. Elle est **symétrique**, puisque $\text{cov}(x, y) = \text{cov}(y, x)$.
2. Sa **diagonale** contient les variances — c'est encore la même identité $\text{cov}(x, x) = V(x)$.

Ce qu'il faut retenir

Cinq points

1. La covariance est la **moyenne du produit des écarts** aux deux moyennes.
2. Elle **généralise la variance** : $\text{cov}(x, x) = V(x)$.
3. Sur le fil rouge, $\text{cov}(x, y) = 72/5 = 14,4$ — positive.
4. Son **signe** s'interprète; sa **valeur**, non : elle dépend des unités et n'est pas bornée.
5. La **matrice des variances-covariances** est symétrique et porte les variances sur sa diagonale.

La suite

Il reste à rendre ce nombre lisible. Le chapitre 5 le divise par les écarts-types — et obtient une mesure bornée entre -1 et $+1$.