

Statistique descriptive

Chapitre 13 : La covariance



Jaouad Madkour

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

D'où vient la formule

Ce que la covariance mesure

Le calcul

Les propriétés

Un cas particulier trompeur

Ce qu'il faut retenir

D'où vient la formule

Rappel du chapitre 8

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

Réécrire le carré comme un produit

Un carré est un produit d'un nombre par lui-même. Écrivons-le ainsi :

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

La variance mesure donc comment x s'écarte de sa moyenne — **en le multipliant par son propre écart.**

La covariance : on ne change qu'une chose

Le second facteur devient l'écart de **l'autre** variable :

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

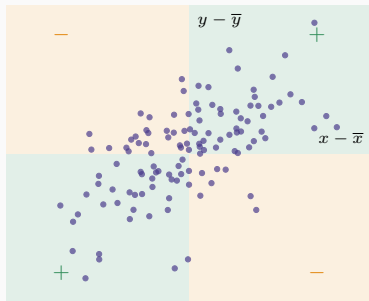
La conséquence immédiate

$$\text{cov}(x, x) = V(x)$$

La **variance** est la covariance d'une variable avec elle-même. Ce n'est pas une analogie : c'est une identité, et elle explique pourquoi les deux notions partagent toutes leurs propriétés.

Ce que la covariance mesure

pourquoi la covariance a le signe de la liaison



D'où vient la formule. La variance mesure comment x s'écarte de sa moyenne, en multipliant cet écart *par lui-même* :

$$V(x) = \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})$$

La covariance ne change qu'une chose : le second facteur devient l'écart de *l'autre* variable.

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

La variance est la covariance d'une variable avec elle-même : $\text{cov}(x, x) = V(x)$.

Chaque point apporte un produit positif s'il est dans un quadrant vert, négatif s'il est dans un quadrant orange. Le signe de la somme dit dans quel sens la liaison penche.

Le raisonnement par quadrants

La contribution d'un individu

Chaque individu apporte à la somme le produit $(x_i - \bar{x})(y_i - \bar{y})$.

Position de l'individu	$x_i - \bar{x}$	$y_i - \bar{y}$	Produit
au-dessus des deux moyennes	+	+	+
au-dessous des deux	—	—	+
au-dessus de x , au-dessous de y	+	—	—
au-dessous de x , au-dessus de y	—	+	—

La lecture

Un individu qui s'écarte **dans le même sens** sur les deux variables apporte du positif; un individu qui s'écarte **en sens contraire** apporte du négatif.

La covariance est le **solde** de ces contributions. Son signe dit dans quel sens penche la liaison.

L'interprétation du signe

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

Le calcul

König-Huygens, version bivariée

$$\text{cov}(x, y) = \frac{1}{n} \sum x_i y_i - \bar{x} \bar{y} = \overline{xy} - \bar{x} \bar{y}$$

Moyenne du produit moins produit des moyennes.

La démonstration, en trois lignes

$$\begin{aligned} \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) &= \frac{1}{n} \sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \overline{xy} - \bar{y} \underbrace{\frac{1}{n} \sum x_i}_{\bar{x}} - \bar{x} \underbrace{\frac{1}{n} \sum y_i}_{\bar{y}} + \bar{x} \bar{y} \\ &= \overline{xy} - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y} = \overline{xy} - \bar{x} \bar{y} \end{aligned}$$

Le parallèle est exact

$$V(x) = \overline{x^2} - \bar{x}^2 \qquad \text{cov}(x, y) = \overline{xy} - \bar{x} \bar{y}$$

Remplacez y par x dans la seconde : vous obtenez la première. **Une seule formule à retenir pour les deux.**

Salaire et ancienneté

$$\overline{xy} = 82\,459 \quad \bar{x} = 10,142 \text{ ans} \quad \bar{y} = 7\,158,26 \text{ dh}$$

$$\text{cov} = 82\,459 - 10,142 \times 7\,158,26 = 82\,459 - 72\,599 = 9\,860$$

Un nombre qu'on ne peut pas interpréter

9 860 dirhams-années. Est-ce beaucoup ?

Impossible de le dire. Trois raisons :

1. L'**unité** est un produit d'unités, qui ne correspond à rien de connu.
2. La covariance n'a **pas de bornes** : elle peut valoir n'importe quoi.
3. Elle dépend des **échelles** : exprimer le salaire en centimes multiplie la covariance par 100, sans que la liaison ait changé d'un iota.

Le signe est interprétable ; la valeur ne l'est pas. C'est exactement le problème que le chapitre 14 résoudra en normalisant.

Les propriétés

Quatre propriétés

1. **Symétrie** : $\text{cov}(x, y) = \text{cov}(y, x)$.
2. **Cas particulier** : $\text{cov}(x, x) = V(x)$.
3. **Bilinéarité** : $\text{cov}(ax + b, cy + d) = ac \text{cov}(x, y)$.
4. **Inégalité de Cauchy-Schwarz** : $|\text{cov}(x, y)| \leq \sigma_x \sigma_y$.

La propriété 3 en clair

Une constante additive **ne change rien** : recentrer les variables ne modifie pas la covariance.

Un changement d'unité la multiplie. C'est la source du problème d'interprétation, et la propriété 4 en donne la solution : **la covariance est bornée par le produit des écarts-types**. Il suffit donc de diviser par ce produit pour obtenir un nombre compris entre -1 et $+1$.

Le résultat le plus utile du chapitre

$$V(x + y) = V(x) + V(y) + 2 \operatorname{cov}(x, y)$$

La variance n'est pas additive

Contrairement à la moyenne, qui l'est toujours : $\overline{x + y} = \bar{x} + \bar{y}$.

Le terme $2 \operatorname{cov}$ est ce qui fait toute la différence. Il s'annule si et seulement si les deux variables ne sont pas corrélées.

La diversification d'un portefeuille

C'est **la** formule de la gestion de portefeuille, et l'un des résultats les plus importants de toute la finance.

Deux actifs de rentabilités x et y . Le risque du portefeuille est $V(x + y)$.

- $\text{cov} > 0$: les deux actifs montent et descendent ensemble. Le risque total **dépasse** la somme des risques.
- $\text{cov} = 0$: les risques s'additionnent simplement.
- $\text{cov} < 0$: les pertes de l'un compensent les gains de l'autre. Le risque total est **inférieur** à la somme.

La diversification n'est rien d'autre que la recherche de covariances négatives. Markowitz a bâti sur cette seule ligne la théorie moderne du portefeuille, et il a obtenu le prix Nobel pour cela.

Le cas extrême

Si $y = -x$, alors $\text{cov}(x, y) = -V(x)$ et

$$V(x + y) = V(x) + V(x) - 2V(x) = 0$$

Risque nul. C'est le principe de la **couverture** : prendre une position exactement opposée pour annuler l'exposition.

Un cas particulier trompeur

Covariance nulle et indépendance

Une implication à sens unique

$$\text{indépendance} \implies \text{cov}(x, y) = 0$$

La réciproque est fausse.

Le contre-exemple à connaître

Prenons $x \in \{-2, -1, 0, 1, 2\}$ et $y = x^2$, donc $y \in \{4, 1, 0, 1, 4\}$.

$\bar{x} = 0$, $\bar{y} = 2$, et

$$\overline{xy} = \frac{(-2)(4) + (-1)(1) + 0 + 1 + 2(4)}{5} = \frac{-8 - 1 + 0 + 1 + 8}{5} = 0$$

$$\text{cov}(x, y) = 0 - 0 \times 2 = 0$$

Et pourtant y est **entièrement déterminé** par x .

Pourquoi la covariance ne le voit pas

Les contributions des quadrants se compensent exactement : pour x négatif, y est au-dessus de sa moyenne dans un quadrant; pour x positif, dans l'autre. Le solde est nul.

La covariance ne détecte que les liaisons monotones et linéaires. Une relation en U lui est invisible.

La conséquence pratique

En finance, deux actifs de covariance nulle ne sont **pas** indépendants pour autant. Ils peuvent parfaitement s'effondrer ensemble lors d'un krach, tout en n'ayant aucune liaison linéaire en temps normal.

C'est l'une des raisons pour lesquelles les modèles de risque fondés sur la seule matrice de covariance ont échoué en 2008 : la dépendance de queue n'y figure pas.

Ce qu'il faut retenir

Six points

1. La covariance s'obtient en remplaçant, dans la variance, le second facteur par l'écart de **l'autre** variable.
2. $\text{cov}(x, x) = V(x)$: la **variance est un cas particulier de covariance**.
3. Formule pratique : $\text{cov}(x, y) = \overline{xy} - \bar{x} \bar{y}$ — moyenne du produit moins produit des moyennes.
4. Son **signe** s'interprète; sa **valeur** non, car elle dépend des unités.
5. $V(x + y) = V(x) + V(y) + 2 \text{cov}(x, y)$: le fondement de la **diversification**.
6. Covariance nulle **n'implique pas** indépendance : la liaison peut être non linéaire.

La suite

Le seul défaut de la covariance est son unité. Cauchy-Schwarz nous a donné la solution : diviser par le produit des écarts-types.

On obtient alors un nombre compris entre -1 et $+1$, comparable d'une étude à l'autre — le **coefficient de corrélation linéaire**. C'est le chapitre 14, et c'est l'aboutissement du cours.