

Statistique descriptive

Chapitre 12 : La covariance

Jaouad Madkour

12 octobre 2026

Plan du chapitre

D'où vient la formule	2
Partir de la variance	2
La covariance : on ne change qu'une chose	2
Ce que la covariance mesure	3
Le mécanisme	3
Le raisonnement par quadrants	3
Le signe de la covariance	3
Des rectangles d'aire signée	4
Le calcul	4
Le calcul par la définition	4
La formule pratique	5
Les deux formules, une seule idée	5
Sur nos 500 salariés	5
Sur un tableau à double entrée	6
Les propriétés	6
Les règles de calcul	6
Les propriétés, en chiffres	6
La variance d'une somme	7
La diversification d'un portefeuille	7
Le cas extrême : la couverture	7
Un cas particulier trompeur	8
Covariance nulle et indépendance	8
Ce que la covariance ne voit pas	8
S'entraîner	8
Exercice 1 : un tableau à double entrée	8
Exercice 1 : solution	9
Exercice 2 : la variance d'un écart de prix	9
Exercice 2 : solution	9

Ce qu'il faut retenir	9
Le résumé du chapitre	9

D'où vient la formule

Partir de la variance

Rappel du chapitre 7

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

Réécrire le carré comme un produit

Un carré est un produit d'un nombre par lui-même :

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

La variance mesure 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

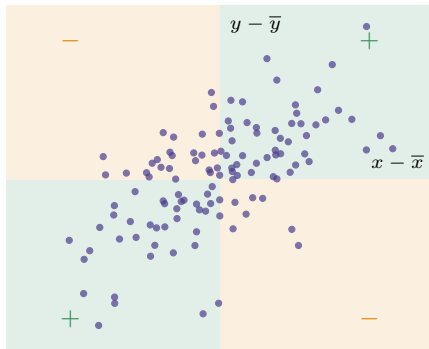
$$\text{cov}(x, x) = V(x)$$

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

Le mécanisme

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 par rapport aux moyennes	$x_i - \bar{x}$	$y_i - \bar{y}$	Produit
au-dessus des deux	+	+	+
au-dessous des deux	-	-	+
x au-dessus, y au-dessous	+	-	-
x au-dessous, y au-dessus	-	+	-

La lecture

Un individu qui s'écarte **dans le même sens** sur les deux variables apporte du positif ; **en sens contraire**, du négatif. La covariance est le **solde** de ces contributions.

Le signe de la covariance

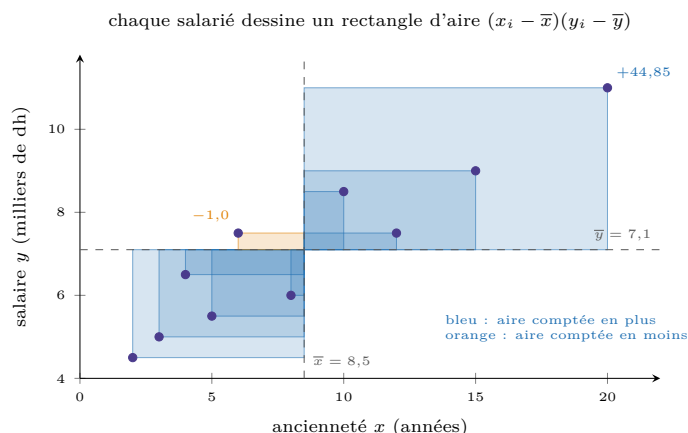
L'interprétation du signe

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

En économie

Salaire et ancienneté, revenu et consommation : covariance positive. Prix et quantité demandée, taux d'intérêt et investissement : covariance négative.

Des rectangles d'aire signée



La covariance comme aire moyenne

Chaque produit est l'**aire** d'un rectangle tendu entre le point moyen et l'individu, comptée en plus ou en moins : la covariance est leur **aire moyenne**. Ici, les dix salariés de l'exercice du chapitre 10.

Le calcul

Le calcul par la définition

Les dix salariés : ancienneté x (ans), salaire y (milliers de dh)

x_i	2	3	4	5	6	8	10	12	15	20
$x_i - 8,5$	-6,5	-5,5	-4,5	-3,5	-2,5	-0,5	1,5	3,5	6,5	11,5
$y_i - 7,1$	-2,6	-2,1	-0,6	-1,6	0,4	-1,1	1,4	0,4	1,9	3,9
Produit	16,9	11,55	2,7	5,6	-1,0	0,55	2,1	1,4	12,35	44,85

Le résultat

Somme des produits : 97,0. Donc $\text{cov}(x, y) = 97,0/10 = 9,7$ années $\times$ milliers de dirhams. Un seul produit est négatif ; le dernier salarié pèse à lui seul près de la moitié de la somme.

La formule pratique

König-Huygens, version bivariable

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

Moyenne du produit moins produit des moyennes.

La démonstration, en trois lignes

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

Les deux formules, une seule idée

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.**

Vérification sur les dix salariés

$\sum x_i y_i = 700,5$, donc $\overline{xy} = 70,05$ et

$$\text{cov} = 70,05 - 8,5 \times 7,1 = 70,05 - 60,35 = 9,70 \quad \checkmark$$

Sur nos 500 salariés

Salaire et ancienneté

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

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

Un nombre qu'on ne peut pas interpréter

9860 **dirhams-années** : est-ce beaucoup ? Impossible à dire. L'**unité** est un produit d'unités ; la covariance n'a **pas de bornes** ; elle dépend des **échelles** — le salaire en centimes la multiplierait par 100, sans que la liaison change.

Le signe est interprétable ; la valeur ne l'est pas. Le chapitre 13 résoudra ce problème en normalisant.

Sur un tableau à double entrée

La formule pondérée

Avec les centres x_i et y_j des classes et les effectifs conjoints n_{ij} :

$$\text{cov}(x, y) = \frac{1}{n} \sum_i \sum_j n_{ij} x_i y_j - \bar{x} \bar{y}$$

Le tableau ancienneté $\times$ salaire du chapitre 10

Centres d'ancienneté 2,5 ; 7,5 ; 12,5 ; 17,5 ; 28 ans. Centres de salaire 3 ; 5 ; 7 ; 9 ; 12 ; 17 milliers. Les 30 produits pondérés totalisent 43 894 :

$$\text{cov} = \frac{43\,894}{500} - 10,743 \times 7,21 = 87,79 - 77,46 = 10,33$$

Soit 10 330 dirhams-années, contre 9 860 sur les données exactes.

Les propriétés

Les règles de calcul

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** ; un changement d'unité multiplie la covariance. La propriété 4 donne la solution : **la covariance est bornée par le produit des écarts-types**. Il suffit de diviser par ce produit pour obtenir un nombre compris entre -1 et $+1$.

Les propriétés, en chiffres

Sur les dix salariés, $\text{cov} = 9,7$ années $\times$ milliers de dh

Transformation	Nouvelle covariance
Salaire en dirhams ($\times 1\,000$)	9 700
Ancienneté en mois ($\times 12$)	116,4
Prime de 500 dh pour tous	9,7
Les deux changements d'unité	116 400

Cauchy-Schwarz, vérifiée

$\sigma_x = 5,48$ ans et $\sigma_y = 1,91$ millier, donc $\sigma_x \sigma_y = 10,46$: la covariance, 9,7, est bien inférieure — et pas de beaucoup. Le rapport des deux, 0,93, sera le coefficient de corrélation.

La variance d'une somme

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 $2 \operatorname{cov}$ fait toute la différence ; il s'annule si et seulement si les deux variables ne sont pas corrélées.

La diversification d'un portefeuille

La formule de la gestion de portefeuille

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

- $\operatorname{cov} > 0$: les deux actifs montent et descendent ensemble ; le risque est **amplifié**.
- $\operatorname{cov} = 0$: les variances s'additionnent simplement.
- $\operatorname{cov} < 0$: les pertes de l'un compensent les gains de l'autre ; le risque est **réduit**.

La diversification est la recherche de covariances faibles ou négatives. Markowitz a bâti sur cette ligne la théorie moderne du portefeuille.

Deux actifs de variance 100 (écart-type 10)

$\operatorname{cov}(x, y)$	+60	0	-60	-100
$V(x + y)$	320	200	80	0
Écart-type de $x + y$	17,9	14,1	8,9	0

Le cas extrême : la couverture

Une position exactement opposée

Si $y = -x$, alors $\operatorname{cov}(x, y) = -V(x)$ 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 — la dernière colonne du tableau précédent.

Et la somme des écarts-types ?

Sans covariance, on aurait pu croire que le risque de $x + y$ vaut $10 + 10 = 20$. C'est le **maximum**, atteint seulement si $\operatorname{cov} = \sigma_x \sigma_y$. Dans tous les autres cas, combiner réduit le risque.

Un cas particulier trompeur

Covariance nulle et indépendance

Une implication à sens unique

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 $y \in \{4, 1, 0, 1, 4\}$. Alors $\bar{x} = 0$, $\bar{y} = 2$, et

$$\overline{xy} = \frac{(-2)(4) + (-1)(1) + 0 + 1 + 2(4)}{5} = 0 \quad \text{donc} \quad \text{cov}(x, y) = 0 - 0 \times 2 = 0$$

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

Ce que la covariance ne voit pas

Pourquoi elle 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 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 s'effondrer ensemble lors d'un krach, sans liaison linéaire en temps normal. Les modèles fondés sur la seule matrice de covariance ignorent cette **dépendance dans les queues**.

S'entraîner

Exercice 1 : un tableau à double entrée

Enfants et pièces de 20 logements

Enfants x , pièces y	$y = 2$	$y = 3$	$y = 4$	Total
$x = 0$	4	2	0	6
$x = 1$	2	4	2	8
$x = 2$	0	2	4	6
Total	6	8	6	20

1. Calculer $\bar{x}$ et $\bar{y}$ à partir des marges.
2. Calculer $\sum n_{ij} x_i y_j$, puis la covariance.
3. Calculer $V(x)$ et $V(y)$, et vérifier l'inégalité de Cauchy-Schwarz.

Exercice 1 : solution

Moyennes et covariance

1. $\bar{x} = (0 \times 6 + 1 \times 8 + 2 \times 6)/20 = 1$ enfant ; $\bar{y} = (2 \times 6 + 3 \times 8 + 4 \times 6)/20 = 3$ pièces.
2. La ligne $x = 0$ n'apporte rien. Ligne $x = 1 : 2 \times 2 + 4 \times 3 + 2 \times 4 = 24$. Ligne $x = 2 : 2 \times (2 \times 3 + 4 \times 4) = 44$. Total 68, d'où

$$\text{cov}(x, y) = \frac{68}{20} - 1 \times 3 = 3,4 - 3 = 0,4$$

Cauchy-Schwarz

3. $V(x) = (8 + 24)/20 - 1 = 0,6$ et $V(y) = (24 + 72 + 96)/20 - 9 = 0,6$. Donc $\sigma_x \sigma_y = 0,6$, et $0,4 \leq 0,6$: la liaison est positive — plus d'enfants, plus de pièces — sans être parfaite.

Exercice 2 : la variance d'un écart de prix

Le blé à Fès et à Meknès

Sur une année, le prix hebdomadaire du blé (dirhams par kilo) a un écart-type de 0,30 à Fès (x) et de 0,25 à Meknès (y), avec $\text{cov}(x, y) = 0,06$.

1. Établir la formule de $V(x - y)$ à partir de celle de $V(x + y)$.
2. Calculer l'écart-type de l'écart de prix $x - y$.
3. Un commerçant gagne sa vie sur l'écart entre les deux marchés. Son revenu est-il plus ou moins risqué que le prix de chaque marché ? Pourquoi ?

Exercice 2 : solution

Une somme avec un coefficient négatif

1. $x - y = x + (-y)$, avec $V(-y) = V(y)$ et $\text{cov}(x, -y) = -\text{cov}(x, y)$:

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

2. $V(x - y) = 0,09 + 0,0625 - 0,12 = 0,0325$, d'où un écart-type de 0,18 dirham par kilo.
3. **Moins risqué** : les deux prix évoluent ensemble ($\text{cov} > 0$), et l'écart retranche ce mouvement commun. Une covariance positive, défavorable pour une somme, est favorable pour une différence.

Ce qu'il faut retenir

Le résumé du chapitre

Sept 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 : $\text{cov}(x, y) = \overline{xy} - \bar{x}\bar{y}$; sur un tableau, on pondère par n_{ij} .
4. Son **signe** s'interprète ; sa **valeur** non, car elle dépend des unités.

5. $V(x \pm y) = V(x) + V(y) \pm 2 \text{cov}(x, y)$: le fondement de la **diversification**.
6. $|\text{cov}| \leq \sigma_x \sigma_y$: la clé du chapitre suivant.
7. Covariance nulle **n'implique pas** indépendance : la liaison peut être non linéaire.

La suite

Le seul défaut de la covariance est son unité. Diviser par le produit des écarts-types donne un nombre compris entre -1 et $+1$: le **coefficient de corrélation linéaire**. C'est le chapitre 13, l'aboutissement du cours.