

Statistique descriptive

Chapitre 7 : Les indicateurs de dispersion

Jaouad Madkour

12 octobre 2026

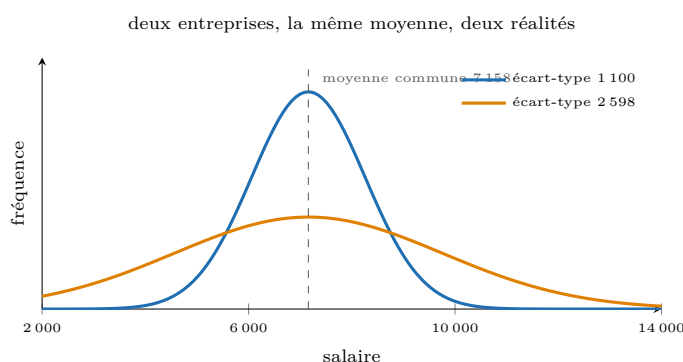
Plan du chapitre

La moyenne ne suffit pas	2
Deux entreprises, une même moyenne	2
Le constat	2
Les mesures simples	3
L'étendue	3
L'écart absolu moyen	3
La variance et l'écart-type	3
Les définitions	3
Des écarts aux carrés des écarts	4
Variance et écart-type des salaires	4
La formule de calcul	4
König-Huygens, en pratique	5
Les propriétés	5
Sur données groupées	5
Les formules pondérées	5
Le tableau de calcul	6
Les trois lignes du calcul	6
Changer d'origine et d'unité	6
Le changement de variable, appliqué	7
Le retour à l'échelle des dirhams	7
Goupé contre exact	7
La dispersion relative	8
Le coefficient de variation	8
Le CV en gestion	8
Le CV par diplôme	8
Un repère de lecture	9
La comparaison des mesures	9
Le tableau récapitulatif	9
Toutes les valeurs, sur nos salaires	9

Le choix	10
S'entraîner	10
Exercice 1 : une variable discrète	10
Exercice 1 : solution	10
Exercice 2 : des loyers en classes	11
Exercice 2 : solution	11
Exercice 2 : les résultats	11
Exercice 3 : une erreur de saisie	12
Exercice 3 : solution	12
Ce qu'il faut retenir	12
Le résumé du chapitre	12

La moyenne ne suffit pas

Deux entreprises, une même moyenne



la moyenne seule ne dit rien
de ce que gagne un salarié
pris au hasard.

**Une moyenne sans dispersion
n'est pas une information.**

$CV = \frac{\sigma}{\bar{x}}$ permet de comparer
deux dispersions exprimées
dans des unités différentes.

Le constat

Ce qu'on cherche

Un **indicateur de dispersion** mesure à quel point les observations s'écartent les unes des autres, ou de leur centre.
Il répond à : la moyenne est-elle représentative ?

La conséquence pratique

Deux entreprises annoncent un salaire moyen de 7 158 dirhams. Dans la première, presque tout le monde gagne entre 6 000 et 8 000. Dans la seconde, les salaires vont de 2 800 à 19 000. Un candidat n'y court pas le même risque.
Une moyenne publiée sans mesure de dispersion n'est pas une information — c'est au mieux un ordre de grandeur.

Les mesures simples

L'étendue

Étendue

$$E = x_{\max} - x_{\min} \quad \text{sur nos salaires : } E = 19\,050 - 2\,810 = 16\,240 \text{ dirhams}$$

Simple, et presque inutilisable

Elle ne dépend que de **deux** observations sur 500 — les deux plus extrêmes, c'est-à-dire les moins représentatives.

Pire : elle **croît mécaniquement** avec n . Plus on observe d'individus, plus on a de chances de rencontrer une valeur extrême. Comparer les étendues de deux échantillons de tailles différentes n'a donc aucun sens. On la mentionne parce qu'elle est intuitive, jamais pour conclure.

L'écart absolu moyen

Écart absolu moyen

$$e = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad \text{sur nos salaires : } e = 1\,972 \text{ dirhams}$$

Un salarié s'écarte en moyenne de 1 972 dirhams du salaire moyen.

Le plus naturel, et presque jamais employé

Il répond exactement à la question posée. Mais la **valeur absolue ne se dérive pas** et ne se manipule pas algébriquement : impossible de la décomposer par sous-groupes ou de la faire entrer dans un intervalle de confiance.

D'où le recours au **carré**, qui remplit la même fonction — annuler les signes — tout en restant dérivable.

La variance et l'écart-type

Les définitions

Variance

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

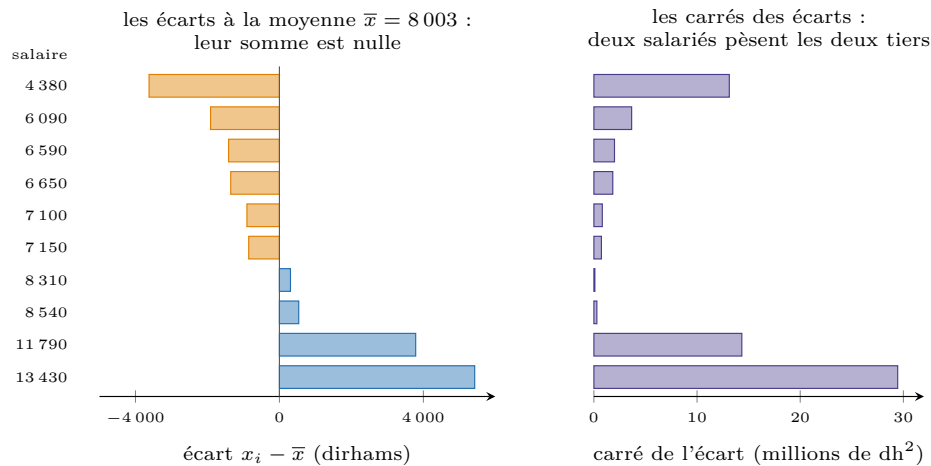
C'est la **moyenne des carrés des écarts à la moyenne**.

Écart-type

$$\sigma_x = \sqrt{V(x)}$$

On prend la racine pour revenir à l'**unité de la variable**.

Des écarts aux carrés des écarts



Les dix premiers salariés du fichier

$\sum (x_i - \bar{x})^2 = 66,33$ millions, donc $V = 6,633$ millions de dh² et $\sigma = 2\,576$ dirhams.

Variance et écart-type des salaires

Sur nos 500 salaires

$$V = 6\,750\,122 \text{ dirhams}^2 \quad \sigma = 2\,598 \text{ dirhams}$$

Pourquoi l'écart-type et pas la variance

La variance s'exprime en **dirhams au carré**, une unité qui ne veut rien dire. On ne peut ni la comparer à la moyenne, ni la reporter sur un graphique.

L'écart-type est en dirhams : il se compare directement à la moyenne, se lit sur l'axe, s'ajoute et se retranche.

On calcule avec la variance, on communique avec l'écart-type.

La formule de calcul

Théorème de König-Huygens

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

La moyenne des carrés moins le carré de la moyenne.

La démonstration, en trois lignes

$$\begin{aligned} \frac{1}{n} \sum (x_i - \bar{x})^2 &= \frac{1}{n} \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \frac{1}{n} \sum x_i^2 - 2\bar{x} \underbrace{\frac{1}{n} \sum x_i}_{=\bar{x}} + \bar{x}^2 = \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 = \overline{x^2} - \bar{x}^2 \end{aligned}$$

König-Huygens, en pratique

Vérification numérique

$$\overline{x^2} = 57\,990\,808 \quad \bar{x}^2 = 7\,158,26^2 = 51\,240\,686$$

$$V = 57\,990\,808 - 51\,240\,686 = 6\,750\,122 \quad \checkmark$$

Pourquoi cette formule compte, au-delà du calcul

Elle ne demande qu'un **seul passage** sur les données : on cumule $\sum x_i$ et $\sum x_i^2$ en même temps. C'est la formule des tableurs et des calculatrices — et c'est celle dont **la covariance sera la généralisation** au chapitre 12. Retenez sa structure : *moyenne du produit moins produit des moyennes*.

Les propriétés

Trois propriétés à connaître

1. $V(x) \geq 0$, et $V(x) = 0$ si et seulement si toutes les valeurs sont égales.
2. $V(x + b) = V(x)$: **ajouter une constante ne change pas la dispersion**.
3. $V(ax) = a^2 V(x)$, donc $\sigma_{ax} = |a| \sigma_x$.

La propriété 2 en clair

Une augmentation générale de 500 dirhams pour tous **ne réduit pas les inégalités** : l'écart-type est inchangé, 2 598 dirhams.

Une augmentation de 10 % pour tous multiplie l'écart-type par 1,1 : il passe à 2 858 dirhams. Elle **augmente** l'écart absolu entre les salariés.

Sur données groupées

Les formules pondérées

Variance d'une distribution groupée

Avec k classes de centres c_i , d'effectifs n_i et de fréquences $f_i = n_i/n$:

$$V(x) = \frac{1}{n} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i c_i^2 - \bar{x}^2 = \sum_{i=1}^k f_i c_i^2 - \bar{x}^2$$

Pour une variable discrète, la valeur x_i remplace le centre c_i .

Rien de nouveau

C'est la définition du début du chapitre, où chaque centre est répété n_i fois : les 500 termes de la somme se regroupent en $k = 6$ paquets. König-Huygens s'applique tel quel.

Le tableau de calcul

Les salaires en six classes, en milliers de dirhams

Classe	c_i	n_i	$n_i c_i$	$n_i c_i^2$
$[2; 4[$	3	31	93	279
$[4; 6[$	5	157	785	3 925
$[6; 8[$	7	166	1 162	8 134
$[8; 10[$	9	79	711	6 399
$[10; 14[$	12	57	684	8 208
$[14; 20[$	17	10	170	2 890
Total		500	3 605	29 835

Pourquoi compter en milliers

En dirhams, la dernière colonne monterait à des milliards : illisible, et source d'erreurs. Diviser par 1 000 divise $\bar{x}$ par 1 000 et V par 1 000² — la propriété $V(ax) = a^2 V(x)$, mise au service du calcul.

Les trois lignes du calcul

Moyenne, moyenne des carrés, variance

$$\bar{x} = \frac{3\,605}{500} = 7,21 \qquad \overline{x^2} = \frac{29\,835}{500} = 59,67$$

$$V(x) = 59,67 - 7,21^2 = 59,67 - 51,984 = 7,686 \quad (\text{milliers de dirhams})^2$$

$$\sigma_x = \sqrt{7,686} = 2,772 \text{ milliers, soit } 2\,772 \text{ dirhams}$$

Une vérification à faire

On peut refaire le calcul par la définition, $\frac{1}{n} \sum n_i (c_i - 7,21)^2$: on trouve le même 7,686. Les deux formules sont équivalentes sur données groupées comme sur données brutes.

Changer d'origine et d'unité

Le changement de variable

On pose $u_i = \frac{c_i - x_0}{a}$, où x_0 est un centre situé vers le milieu du tableau et a une amplitude commode. Alors

$$\bar{x} = x_0 + a \bar{u} \qquad V(x) = a^2 V(u)$$

Ce que cela achète

Les u_i sont de petits nombres, souvent négatifs : les produits $n_i u_i$ et $n_i u_i^2$ se font de tête. Les deux formules ne sont que les propriétés de la moyenne (chapitre 5) et de la variance, lues à l'envers.

Le changement de variable, appliqué

Avec $x_0 = 7000$ et $a = 1000$

c_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$
3 000	-4	31	-124	496
5 000	-2	157	-314	628
7 000	0	166	0	0
9 000	2	79	158	316
12 000	5	57	285	1 425
17 000	10	10	100	1 000
Total		500	105	3 865

Le retour à l'échelle des dirhams

Le résultat

$$\bar{u} = \frac{105}{500} = 0,21 \quad V(u) = \frac{3865}{500} - 0,21^2 = 7,73 - 0,0441 = 7,6859$$
$$\bar{x} = 7000 + 1000 \times 0,21 = 7210 \quad V(x) = 1000^2 \times 7,6859 = 7685900$$

On retrouve $\sigma_x = 2772$ dirhams, avec des calculs qui tiennent sur une feuille.

Le choix de x_0 et de a

N'importe quel x_0 et n'importe quel a conviennent : le résultat est exact quel que soit le choix. On prend un centre proche du milieu, pour que les u_i soient petits, et l'amplitude la plus fréquente, pour qu'ils soient entiers.

Groupe contre exact

Ce que coûte le regroupement

Indicateur	Groupe	Exact	Écart
Moyenne	7 210	7 158	+0,7 %
Écart absolu moyen	2 049	1 972	+3,9 %
Écart-type	2 772	2 598	+6,7 %
Variance	7 685 900	6 750 122	+13,9 %
CV	0,385	0,363	

Pourquoi le regroupement exagère ici la dispersion

Les centres poussent les classes extrêmes **vers l'extérieur** (chapitre 3) : dans $[2000; 4000[$, la moyenne réelle est 3 528, pas 3 000 ; dans $[14000; 20000[$, 15 927, pas 17 000. Les écarts sont gonflés aux deux bouts, et le carré amplifie le gonflement. **Une variance tirée d'un tableau en classes se donne avec deux ou trois chiffres significatifs.**

La dispersion relative

Le coefficient de variation

Coefficient de variation

$$CV = \frac{\sigma_x}{\bar{x}}$$

C'est un nombre **sans unité**, souvent exprimé en pourcentage. Il exige $\bar{x} > 0$.

Comparer deux variables incomparables

Variable	Moyenne	Écart-type	CV
Salaire	7 158 dh	2 598 dh	0,363
Ancienneté	10,14 ans	5,96 ans	0,588

Les écarts-types, l'un en dirhams et l'autre en années, ne se comparent pas. Les *CV*, si : **l'ancienneté est nettement plus dispersée que le salaire**, relativement à son niveau.

Le CV en gestion

Deux usages

1. **Comparer des variables d'unités différentes**, comme ci-dessus.
2. **Comparer des populations de niveaux très différents** : la dispersion des salaires d'une multinationale et d'une PME.

En finance, le *CV* s'inverse et porte un autre nom : le rapport de la rentabilité moyenne à sa volatilité est le **ratio de Sharpe**. Même idée — rapporter la dispersion au niveau — au service d'une autre question.

Le CV par diplôme

Trois groupes, trois niveaux

Diplôme	Moyenne	Écart-type	CV
Bac	6 059	2 021	0,334
Bac+2	7 081	2 045	0,289
Bac+5	9 534	2 872	0,301

En dirhams, les Bac+5 sont les plus dispersés ; **relativement à leur niveau**, ce sont les Bac.

Deux questions différentes

« Quel groupe a les écarts de salaire les plus grands ? » : les Bac+5, $\sigma = 2\,872$ dirhams. « Quel groupe est le moins homogène ? » : les Bac, $CV = 0,33$. L'écart-type répond à la première, le *CV* à la seconde.

Un repère de lecture

Ordres de grandeur

CV	Lecture usuelle
$< 0,15$	population très homogène
$0,15 \text{ à } 0,30$	dispersion modérée
$0,30 \text{ à } 0,60$	dispersion forte
$> 0,60$	population très hétérogène

Ces seuils sont indicatifs, pas normatifs

Ils dépendent entièrement du domaine. Un CV de 0,05 serait énorme pour un processus industriel sous contrôle, et dérisoire pour des rentabilités boursières. **Le bon repère est toujours la comparaison** : le même indicateur, sur une autre population ou une autre période.

La comparaison des mesures

Le tableau récapitulatif

Les six indicateurs de dispersion

Indicateur	Formule	Robuste ?	Unité
Étendue	$x_{\max} - x_{\min}$	non	celle de x
Intervalle interquartile	$Q_3 - Q_1$	oui	celle de x
Écart absolu moyen	$\frac{1}{n} \sum x_i - \bar{x} $	assez	celle de x
Variance	$\overline{x^2} - \bar{x}^2$	non	celle de x au carré
Écart-type	$\sqrt{V}$	non	celle de x
Coefficient de variation	$\sigma/\bar{x}$	non	aucune

Toutes les valeurs, sur nos salaires

Cinq indicateurs, une même distribution

Indicateur	Valeur
Étendue	16 240
IQ	3 012
Écart absolu moyen	1 972
Écart-type	2 598

CV

0,363

Deux remarques sur ces nombres

L'écart-type dépasse l'écart absolu moyen : 2 598 contre 1 972. C'est toujours le cas, parce que le carré donne plus de poids aux grands écarts. Plus la distribution a une queue longue, plus l'écart-type s'éloigne de l'écart absolu moyen.

Le choix**En pratique**

- Pour **décrire** une distribution asymétrique ou comportant des valeurs extrêmes : IQ et boîte à moustaches.
- Pour **calculer** quoi que ce soit ensuite : variance et écart-type. Ils sont les seuls à entrer dans les formules de la covariance, de la corrélation et de toute l'inférentielle.
- Pour **comparer** des unités différentes : coefficient de variation.

Et dans un tableau de résultats professionnel, on donne les trois.

S'entraîner**Exercice 1 : une variable discrète****Le nombre d'enfants à charge**

Une enquête auprès de 40 salariés donne la répartition suivante.

Nombre d'enfants x_i	0	1	2	3	4
Effectif n_i	6	10	12	8	4

1. Calculer la moyenne $\bar{x}$.
2. Calculer la variance par la formule de König-Huygens, puis l'écart-type.
3. Calculer le coefficient de variation et commenter.

Exercice 1 : solution**Le tableau de calcul**

x_i	n_i	$n_i x_i$	$n_i x_i^2$
0	6	0	0
1	10	10	10
2	12	24	48
3	8	24	72
4	4	16	64
Total	40	74	194

Les résultats

1. $\bar{x} = 74/40 = 1,85$ enfant.
2. $\overline{x^2} = 194/40 = 4,85$, d'où $V(x) = 4,85 - 1,85^2 = 1,4275$, et $\sigma_x = 1,19$ enfant.
3. $CV = 1,19/1,85 = 0,65$: population **très hétérogène** au sens du repère du chapitre.

Exercice 2 : des loyers en classes

Les loyers de 100 ménages

Loyer (dirhams)	n_i
[1 000 ; 1 500[	8
[1 500 ; 2 000[	20
[2 000 ; 2 500[	32
[2 500 ; 3 000[	25
[3 000 ; 4 000[	15

1. Calculer les centres c_i , puis $u_i = (c_i - 2\,250)/500$.
2. En déduire $\bar{x}$, $V(x)$ et σ_x par le changement de variable.
3. Calculer le coefficient de variation et le situer sur le repère du chapitre.

Exercice 2 : solution

Le tableau, en variable u

Classe	c_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$
[1 000 ; 1 500[	1 250	-2	8	-16	32
[1 500 ; 2 000[	1 750	-1	20	-20	20
[2 000 ; 2 500[	2 250	0	32	0	0
[2 500 ; 3 000[	2 750	1	25	25	25
[3 000 ; 4 000[	3 500	2,5	15	37,5	93,75
Total			100	26,5	170,75

Exercice 2 : les résultats

Retour à l'échelle des dirhams

$$\bar{u} = \frac{26,5}{100} = 0,265 \quad V(u) = \frac{170,75}{100} - 0,265^2 = 1,7075 - 0,0702 = 1,6373$$

$$\bar{x} = 2\,250 + 500 \times 0,265 = 2\,382,5 \text{ dh} \quad V(x) = 500^2 \times 1,6373 = 409\,319 \text{ dh}^2$$

$$\sigma_x = \sqrt{409\,319} = 640 \text{ dh} \quad CV = \frac{640}{2\,382,5} = 0,27$$

La lecture

Dispersion **modérée** (entre 0,15 et 0,30). La dernière classe, deux fois plus large, donne $u = 2,5$: le changement de variable n'exige pas des amplitudes égales.

Exercice 3 : une erreur de saisie

Corriger sans tout recalculer

Pour 20 salariés, un logiciel a calculé $\sum x_i = 160\,000$ et $\sum x_i^2 = 1\,324\,000\,000$ (salaires en dirhams).

1. Calculer $\bar{x}$, $V(x)$ et σ_x .
2. On découvre qu'un salaire de 9 000 a été saisi 6 000. Corriger $\sum x_i$ et $\sum x_i^2$, puis recalculer $\bar{x}$ et σ_x .
3. L'erreur gonflait-elle ou réduisait-elle la dispersion ? Pourquoi ?

Exercice 3 : solution

Les sommes suffisent

1. $\bar{x} = 160\,000/20 = 8\,000$; $\overline{x^2} = 66\,200\,000$; $V = 66\,200\,000 - 64\,000\,000 = 2\,200\,000$; $\sigma_x = 1\,483$ dh.
2. $\sum x_i = 160\,000 - 6\,000 + 9\,000 = 163\,000$ et $\sum x_i^2 = 1\,324\,000\,000 - 36\,000\,000 + 81\,000\,000 = 1\,369\,000\,000$. D'où $\bar{x} = 8\,150$, $\overline{x^2} = 68\,450\,000$, $V = 68\,450\,000 - 66\,422\,500 = 2\,027\,500$ et $\sigma_x = 1\,424$ dh.
3. L'erreur **gonflait** la dispersion : la valeur fautive, 6 000, était à 2 000 dh sous la moyenne ; la vraie, 9 000, n'est qu'à 850 dh au-dessus de la moyenne corrigée.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. Une moyenne sans dispersion n'est pas une information.
2. L'**étendue** ne repose que sur deux observations et croît avec n : à éviter.
3. **Variance** : moyenne des carrés des écarts. **Écart-type** : sa racine, dans l'unité de la variable.
4. **König-Huygens** : $V(x) = \overline{x^2} - \bar{x}^2$ — moyenne du produit moins produit des moyennes.
5. $V(x + b) = V(x)$ et $V(ax) = a^2 V(x)$: une prime uniforme ne réduit pas les écarts.
6. **Données groupées** : $V(x) = \sum f_i c_i^2 - \bar{x}^2$; le changement de variable $u = (c - x_0)/a$ allège le calcul. Le résultat reste une approximation.
7. **Coefficient de variation** $CV = \sigma/\bar{x}$: sans unité, il permet de comparer l'incomparable.

La suite

Deux distributions peuvent avoir la même moyenne **et** le même écart-type sans se ressembler : l'une symétrique, l'autre étalée d'un côté. Il faut aller plus loin — au moment d'ordre trois, puis d'ordre quatre. C'est le chapitre 8.