

Statistique descriptive

Chapitre 5 : Les indicateurs de tendance centrale

Jaouad Madkour

12 octobre 2026

Plan du chapitre

La question	2
Un seul nombre pour une distribution	2
Le mode	2
Le mode : la définition	2
Le mode sur nos données	3
Préciser le mode dans la classe modale	3
La construction graphique	3
Les défauts du mode	4
Une bimodalité est une information	4
La médiane	4
La médiane : la définition	4
La médiane sur nos données	5
Le calcul sur données groupées	5
La médiane est robuste	5
La moyenne arithmétique	6
La moyenne : la définition	6
La moyenne sur données groupées	6
Somme des écarts et linéarité	6
La moyenne des sous-groupes	7
Ce que minimisent la moyenne et la médiane	7
Le défaut de la moyenne	7
Les autres moyennes	8
Trois moyennes de plus	8
Les quatre moyennes des salaires	8
L'inégalité des moyennes	8
La géométrique : les taux de croissance	9
L'harmonique et la quadratique	9
Choisir	9
Le tableau de décision	9

Moyenne pour agréger, médiane pour décrire	10
Le diagnostic par comparaison	10
La règle empirique	10
Une conséquence à retenir	11
S'entraîner	11
Exercice 1 : les trois indicateurs sur données groupées	11
Exercice 1 : solution	11
Exercice 2 : une variable discrète	12
Exercice 2 : solution	12
Ce qu'il faut retenir	12
Le résumé du chapitre	12

La question

Un seul nombre pour une distribution

Ce qu'on cherche

Un **indicateur de tendance centrale** est une valeur unique qui résume « où se situe » l'ensemble des observations.

Il répond à : autour de quoi les données se rassemblent-elles ?

Trois réponses, trois définitions du mot central

- Le **mode** : la valeur la plus **fréquente**.
- La **médiane** : la valeur qui **partage** l'effectif en deux.
- La **moyenne** : la valeur qui **équilibre** la somme.

Ces trois réponses ne coïncident que si la distribution est symétrique. Dès qu'elle est asymétrique, elles divergent — et c'est précisément l'écart entre elles qui devient informatif.

Le mode

Le mode : la définition

Mode

Le **mode** est la modalité de **plus grand effectif**.

Pour une variable continue, on parle de **classe modale** : la classe de plus grande **densité** — pas de plus grand effectif.

Le seul indicateur disponible pour une variable nominale

Quelle est la région « moyenne » ? La question n'a aucun sens. Quelle est la région la plus représentée ? Tanger-Tétouan, avec 36 % des salariés.

Le mode est le seul résumé possible d'une variable nominale — et c'est pour cela qu'il figure en premier.

Le mode sur nos données

Une variable discrète : l'ancienneté

Le mode se lit directement dans le tableau des valeurs : l'ancienneté la plus fréquente est de **6 ans**, avec 47 salariés, devant 4 ans (39 salariés) et 10 ans (38).

Une variable continue : la classe modale des salaires

La densité est maximale dans $[6\,000; 8\,000[$: $d = 0,0830$, soit 83 salariés par tranche de 1 000 dirhams.

C'est bien la classe modale, alors que la classe $[4\,000; 6\,000[$ la talonne : 157 salariés contre 166, avec la même amplitude.

Préciser le mode dans la classe modale

Le mode par interpolation

Avec Δ_1 l'écart de densité avec la classe **précédente** et Δ_2 avec la classe **suivante** :

$$Mo = e_{i-1} + a_i \times \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

Le calcul sur les salaires

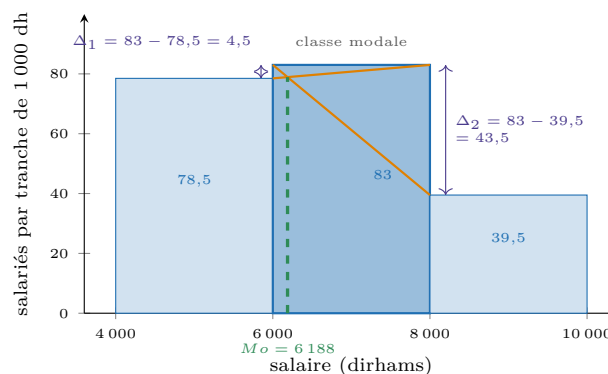
Densités, en salariés par tranche de 1 000 dirhams : 78,5, 83, 39,5. Donc $\Delta_1 = 4,5$ et $\Delta_2 = 43,5$:

$$Mo = 6\,000 + 2\,000 \times \frac{4,5}{4,5 + 43,5} = 6\,000 + 2\,000 \times 0,094 = 6\,188 \text{ dirhams}$$

La classe précédente est presque aussi dense : le mode se place **près de la borne basse**.

La construction graphique

la construction graphique du mode



Deux diagonales

On joint chaque coin haut de la classe modale au coin haut de la classe voisine, en croisant. Les deux segments se coupent à l'abscisse du mode : c'est la formule, dessinée.

Les défauts du mode

Trois limites

1. Il peut ne pas être **unique** : une distribution bimodale a deux modes.
2. Il dépend entièrement du **découpage en classes** : changer les bornes change le mode.
3. Il **ignore tout le reste** de la distribution.

Un mode qui dépend du découpage

- classes de 2 000 : $[6\,000; 8\,000[$, avec 166 salariés ;
- classes de 1 000 : $[5\,000; 6\,000[$ (90), juste devant $[7\,000; 8\,000[$ (89) ;
- classes de 500 : égalité, 49 salariés en $[5\,500; 6\,000[$ et en $[7\,000; 7\,500[$.

Une bimodalité est une information

Deux bosses, deux populations

Deux bosses sur un histogramme de salaires signifient presque toujours **deux populations mélangées** : cadres et employés, temps plein et temps partiel, deux établissements. Le bon réflexe n'est alors pas de calculer un indicateur global, mais de **séparer les deux groupes** et de les décrire chacun.

Et ici ?

Avec des classes de 500 dirhams, nos salaires semblent avoir deux sommets. Mais avec 500 salariés, une classe de 500 dirhams ne contient qu'une quarantaine d'individus : l'écart entre deux classes voisines est du **bruit**. Une vraie bimodalité doit résister à un changement de découpage.

La médiane

La médiane : la définition

Médiane

La **médiane** Me est la valeur qui partage la population **en deux parties d'effectifs égaux** : la moitié des individus sont en dessous, la moitié au-dessus.

Le calcul sur données brutes

On ordonne les n valeurs.

- Si n est **impair** : Me est la valeur de rang $\frac{n+1}{2}$.
- Si n est **pair** : Me est la moyenne des valeurs de rangs $\frac{n}{2}$ et $\frac{n}{2} + 1$.

La médiane sur nos données

Le salaire

$n = 500$ est pair. On prend la moyenne des salaires de rangs 250 et 251 :

$$Me = 6\,770 \text{ dirhams}$$

L'ancienneté, avec les effectifs cumulés

On cherche les rangs 250 et 251 dans la colonne N :

Ancienneté	≤ 8 ans	≤ 9 ans
Effectif cumulé N	227	262

Les rangs 228 à 262 ont exactement 9 ans : les rangs 250 et 251 aussi. $Me = 9$ **ans**.

Le calcul sur données groupées

L'interpolation linéaire

On repère la classe où F franchit 0,5, puis :

$$Me = e_{i-1} + a_i \times \frac{0,5 - F_{i-1}}{F_i - F_{i-1}}$$

Le calcul détaillé

F passe de 0,376 à 0,708 dans la classe $[6\,000 ; 8\,000[$.

$$Me = 6\,000 + 2\,000 \times \frac{0,500 - 0,376}{0,708 - 0,376} = 6\,000 + 2\,000 \times 0,3735 = 6\,747$$

Contre 6 770 sur les données exactes : l'erreur d'équirépartition vaut 23 dirhams.

La médiane est robuste

Elle ne dépend que des rangs

La médiane ne dépend que du **rang** des observations, pas de leur valeur.

Remplacer le salaire le plus élevé, 19 050 dirhams, par 190 050 dirhams : la médiane ne bouge **pas d'un dirham**. La moyenne, elle, passe de 7 158 à 7 500.

Pourquoi les revenus se publient en médiane

Les distributions de revenus, de patrimoines, de prix de l'immobilier sont fortement asymétriques : quelques valeurs très élevées tirent la moyenne. Le **salaire médian** décrit une personne réelle : celle qui est au milieu. Ici, 56 % des salariés gagnent **moins** que la moyenne. **C'est pourquoi les institutions statistiques publient la médiane, et pourquoi les communiqués d'entreprise publient la moyenne.**

La moyenne arithmétique

La moyenne : la définition

Moyenne arithmétique

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{ou, sur données groupées,} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k n_i c_i = \sum_{i=1}^k f_i c_i$$

Sur nos données

$$\bar{x} = \frac{3\,579\,130}{500} = 7\,158,26 \text{ dirhams}$$

La moyenne sur données groupées

Avec les fréquences, en milliers de dirhams

Classe	c_i	f_i	$f_i c_i$
$[2; 4[$	3	0,062	0,186
$[4; 6[$	5	0,314	1,570
$[6; 8[$	7	0,332	2,324
$[8; 10[$	9	0,158	1,422
$[10; 14[$	12	0,114	1,368
$[14; 20[$	17	0,020	0,340
Total		1	7,210

Une moyenne pondérée

Chaque centre pèse sa fréquence : $\bar{x} = 7\,210$ dirhams. Les deux classes centrales apportent à elles seules 3,9 des 7,2 milliers.

Somme des écarts et linéarité

Propriété 1 : la somme des écarts est nulle

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

La moyenne est le **point d'équilibre** de la distribution : les écarts positifs compensent exactement les écarts négatifs.

Propriété 2 : linéarité

$$\overline{ax + b} = a\bar{x} + b$$

Une augmentation générale de 500 dirhams augmente la moyenne de 500 dirhams. Une hausse de 10 % la multiplie par 1,1.

La moyenne des sous-groupes

Propriété 3 : la moyenne totale

Si la population est partagée en groupes d'effectifs n_j et de moyennes $\bar{x}_j$:

$$\bar{x} = \frac{1}{n} \sum_j n_j \bar{x}_j$$

C'est le **théorème de la moyenne totale** : une moyenne **pondérée par les effectifs**.

Vérification sur le diplôme

$$\frac{219 \times 6\,059 + 174 \times 7\,081 + 107 \times 9\,534}{500} = 7\,158$$

Attention : la moyenne simple des trois valeurs, $(6\,059 + 7\,081 + 9\,534)/3 = 7\,558$, est **fausse** de 400 dirhams. C'est l'erreur la plus fréquente sur les moyennes.

Ce que minimisent la moyenne et la médiane

Deux façons d'être au centre

- La **moyenne** rend minimale la somme des **carrés** des écarts $\sum (x_i - a)^2$.
- La **médiane** rend minimale la somme des écarts **absolus** $\sum |x_i - a|$.

Sur cinq valeurs : 2, 3, 4, 8, 13

Médiane 4, moyenne $30/5 = 6$. On essaie trois candidats a :

a	$\sum x_i - a $	$\sum (x_i - a)^2$
4 (médiane)	16	102
5	17	87
6 (moyenne)	18	82

Le carré pénalise lourdement le grand écart à 13 : la moyenne s'en rapproche, la médiane l'ignore.

Le défaut de la moyenne

Elle est sensible aux valeurs extrêmes

Elle utilise **toutes** les valeurs, ce qui est sa force pour le calcul et sa faiblesse pour l'interprétation.

Une seule valeur aberrante — une erreur de saisie, un dirigeant au salaire hors norme — la déplace sensiblement. La médiane, elle, ne bronche pas.

Le réflexe professionnel

Avant de publier une moyenne, on regarde les **valeurs extrêmes** : sont-elles réelles ou des erreurs ? Si elles sont réelles, on publie la moyenne **et** la médiane ; leur écart dit déjà quelque chose de la distribution.

Les autres moyennes

Trois moyennes de plus

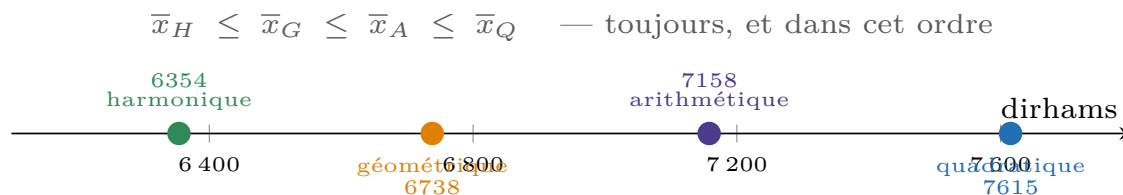
Les définitions

Moyenne	Formule	Quand l'employer
Géométrique	$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}$	des taux de croissance
Harmonique	$\bar{x}_H = \frac{1}{\sum 1/x_i}$	des rapports, des vitesses
Quadratique	$\bar{x}_Q = \sqrt{\frac{1}{n} \sum x_i^2}$	des écarts, des erreurs

Une même idée

Chacune est la moyenne arithmétique d'une **transformation** des données — logarithme, inverse, carré — ramenée ensuite à l'échelle de départ.

Les quatre moyennes des salaires



L'inégalité des moyennes

Un ordre qui ne change jamais

$$\bar{x}_H \leq \bar{x}_G \leq \bar{x}_A \leq \bar{x}_Q$$

avec égalité si et seulement si toutes les valeurs sont identiques. **Sur nos salaires** : $6\,354 \leq 6\,738 \leq 7\,158 \leq 7\,615$.

Ce que l'écart mesure

Plus les quatre moyennes sont **écartées**, plus la distribution est dispersée. Si toutes les valeurs étaient égales, les quatre coïncideraient.

L'écart entre $\bar{x}_G$ et $\bar{x}_A$ est même une mesure classique d'inégalité — c'est le fondement de l'indice d'Atkinson, employé en économie du bien-être.

La géométrie : les taux de croissance

Une hausse, puis une baisse de même taux

Un chiffre d'affaires croît de +20 % une année, puis de −20 % l'année suivante. La croissance moyenne est-elle nulle ?

Non. $100 \rightarrow 120 \rightarrow 96$. La moyenne **arithmétique** des taux donne 0 %, ce qui est faux. Il faut la moyenne **géométrique** des coefficients multiplicateurs :

$$\sqrt{1,20 \times 0,80} = \sqrt{0,96} = 0,9798$$

soit un taux moyen de −2,02 % par an. Ce qui est juste : $100 \times 0,9798^2 = 96$.

L'harmonique et la quadratique

L'harmonique : les rapports

Un commercial parcourt 100 km à 80 km/h à l'aller, 100 km à 120 km/h au retour. Le trajet dure $1,25 + 0,833 = 2,083$ heures pour 200 km, soit 96 km/h — pas 100.

$$\bar{x}_H = \frac{2}{\frac{1}{80} + \frac{1}{120}} = 96$$

Règle : pour moyenner un **rapport** — km/h, dirhams par unité — à **numérateur constant**, c'est l'harmonique.

La quadratique : les écarts

Elle moyenne des grandeurs dont le signe ne doit pas compenser : écarts à une cible, erreurs de prévision. Appliquée aux écarts $x_i - \bar{x}$, elle **est** l'écart-type du chapitre 7.

Choisir

Le tableau de décision

Lequel employer

Situation	Indicateur
Variable nominale	mode, et lui seul
Variable ordinale	mode, médiane
Distribution symétrique	moyenne
Distribution asymétrique	médiane
Valeurs aberrantes possibles	médiane
Il faut pouvoir recomposer un total	moyenne
Taux de croissance	moyenne géométrique
Rapports, vitesses	moyenne harmonique

Moyenne pour agréger, médiane pour décrire

Le seul argument décisif en faveur de la moyenne

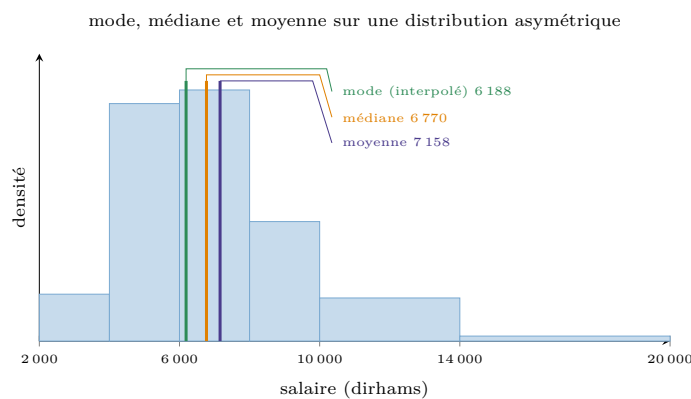
Elle est la seule qui permette de **remonter au total** : $\bar{x} \times n$ redonne la masse salariale, 3 579 130 dirhams.

C'est indispensable en comptabilité, en budget, en agrégation. La médiane, elle, ne s'additionne pas : la médiane de deux populations réunies n'est pas la moyenne des deux médianes.

La vérification sur le fil rouge

Médiane des hommes 6 820, des femmes 6 580 : leur moyenne pondérée vaut $(285 \times 6 820 + 215 \times 6 580)/500 = 6 717$. La médiane de l'ensemble est 6 770. **La médiane ne se recompose pas.**

Le diagnostic par comparaison



distribution étalée à droite :

mode < médiane < moyenne

Les hauts salaires, peu nombreux, tirent la moyenne vers le haut ; la médiane, elle, ne bouge pas.

D'où la règle : pour un revenu, **publier la médiane**. La moyenne décrit une personne qui n'existe pas.

La règle empirique

Le sens de l'écart

Configuration	Forme de la distribution
mode = médiane = moyenne	symétrique
mode < médiane < moyenne	étalée à droite
mode > médiane > moyenne	étalée à gauche

Sur nos salaires

$$6 188 < 6 770 < 7 158$$

Mode, médiane, moyenne dans cet ordre : distribution **étalée à droite**. Ce que l'histogramme montrait déjà, et que le chapitre 8 chiffrera avec le coefficient d'asymétrie.

Une conséquence à retenir

La majorité sous la moyenne

Dans une distribution étalée à droite, la **majorité** des individus gagne moins que la moyenne. Ici 56 %.

« Le salaire moyen est de 7 158 dirhams » est donc une phrase exacte et trompeuse. C'est le genre de nuance qu'un économiste doit savoir signaler.

Et l'ancienneté ?

Mode 6 ans, médiane 9 ans, moyenne 10,1 ans : même ordre, même diagnostic. Quelques salariés très anciens tirent la moyenne vers le haut.

S'entraîner

Exercice 1 : les trois indicateurs sur données groupées

Le chiffre d'affaires journalier d'une boutique

Sur 60 jours, en milliers de dirhams :

Classe	$[0 ; 2[$	$[2 ; 4[$	$[4 ; 6[$	$[6 ; 10[$	$[10 ; 16[$
Jours n_i	6	18	21	12	3

1. Calculer la moyenne.
2. Calculer la médiane par interpolation.
3. Déterminer la classe modale, puis le mode par interpolation. Attention aux amplitudes.
4. Que dit la position des trois indicateurs ?

Exercice 1 : solution

Moyenne et médiane

1. Centres 1, 3, 5, 8, 13 : $\sum n_i c_i = 300$, donc $\bar{x} = 300/60 = 5,0$ milliers de dirhams.
2. F : 0,10 ; 0,40 ; 0,75 ; 0,95 ; 1. On franchit 0,5 dans $[4 ; 6[$:

$$Me = 4 + 2 \times \frac{0,50 - 0,40}{0,75 - 0,40} = 4 + 0,571 = 4,57$$

Mode et diagnostic

3. Densités : 3 ; 9 ; 10,5 ; $12/4 = 3$; $3/6 = 0,5$. Classe modale $[4 ; 6[$, avec $\Delta_1 = 10,5 - 9 = 1,5$ et $\Delta_2 = 10,5 - 3 = 7,5$:

$$Mo = 4 + 2 \times \frac{1,5}{1,5 + 7,5} = 4,33$$

4. $4,33 < 4,57 < 5,0$: distribution **étalée à droite**.

Exercice 2 : une variable discrète

Les jours d'absence de 80 salariés

Jours d'absence x_i	0	1	2	3	4	5	8
Salariés n_i	22	18	15	10	7	5	3

1. Donner le mode.
2. Calculer la médiane à l'aide des effectifs cumulés.
3. Calculer la moyenne.
4. Le directeur annonce « près de deux jours d'absence par salarié ». Quelle part des salariés est effectivement absente au moins deux jours ?

Exercice 2 : solution

Les trois indicateurs

1. Mode : 0 jour (22 salariés).
2. Effectifs cumulés : 22 ; 40 ; 55 ; 65 ; 72 ; 77 ; 80. Avec $n = 80$, on cherche les rangs 40 et 41 : le rang 40 vaut 1, le rang 41 vaut 2. $Me = (1 + 2)/2 = 1,5$ jour.
3. $\sum n_i x_i = 0 + 18 + 30 + 30 + 28 + 25 + 24 = 155$, donc $\bar{x} = 155/80 = 1,94$ jour.

La lecture

4. Au moins deux jours : $80 - 40 = 40$ salariés, soit la **moitié**. L'autre moitié a manqué un jour au plus. $0 < 1,5 < 1,94$: la distribution est étalée à droite, et les trois salariés absents 8 jours pèsent à eux seuls 24 des 155 jours.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. **Mode** : le plus fréquent. Seul indicateur possible pour une variable nominale ; il dépend du découpage.
2. **Médiane** : partage l'effectif en deux. **Robuste** aux valeurs extrêmes.
3. **Moyenne** : point d'équilibre. Linéaire, et se recompose par groupes — **pondérée par les effectifs**.
4. Sur données groupées : moyenne par les centres, médiane et mode **par interpolation**.
5. Géométrique pour les **taux**, harmonique pour les **rapports**, quadratique pour les **écarts**.
6. $\bar{x}_H \leq \bar{x}_G \leq \bar{x}_A \leq \bar{x}_Q$, toujours.
7. La position relative du mode, de la médiane et de la moyenne révèle le **sens de l'asymétrie**.

La suite

La médiane coupe la distribution en deux. Rien n'empêche de la couper en quatre, en dix, en cent — et l'on obtient une description bien plus fine de la répartition. C'est le chapitre 6 : les quantiles.