

Statistique descriptive

Chapitre 9 : La concentration

Jaouad Madkour

12 octobre 2026

Plan du chapitre

Une question nouvelle	2
Dispersion et concentration ne sont pas la même chose	2
Où la question se pose	2
Dispersion contre concentration, en chiffres	2
La courbe de Lorenz	3
La construction	3
Comment on la trace	3
Lire la courbe	4
Lorenz sur données groupées	4
La courbe des six classes	5
L'indice de Gini	5
La définition	5
La formule de calcul	5
Le Gini par les trapèzes	6
Le calcul sur les six classes	6
L'interprétation	6
Ce que la comparaison enseigne	7
Ses limites	7
La médiale	7
Une mesure complémentaire	7
La médiale sur données groupées	8
L'écart entre médiale et médiane	8
Bilan de la partie univariée	8
Les cinq familles	8
La fiche complète de nos salaires	9
S'entraîner	9
Exercice 1 : la concentration des terres	9
Exercice 1 : solution	9
Exercice 2 : prime uniforme ou hausse proportionnelle	10

Exercice 2 : solution	10
Ce qu'il faut retenir	10
Le résumé du chapitre	10

Une question nouvelle

Dispersion et concentration ne sont pas la même chose

La question de la concentration

Comment la **masse totale** se répartit-elle entre les individus ?
Autrement dit : les 10 % les mieux payés reçoivent-ils 10 % de la masse salariale, ou 40 % ?

La différence avec la dispersion

La **dispersion** demande : les individus sont-ils éloignés les uns des autres ? La **concentration** demande : la somme est-elle également partagée ?
Ce n'est pas une nuance. Ajouter 1 000 dirhams à tout le monde ne change **pas** l'écart-type — mais **réduit** la concentration, puisque les écarts relatifs diminuent.

Où la question se pose

Quatre domaines

Domaine	Question
Revenus, patrimoines	quelle part détenue par les plus riches ?
Marché, concurrence	quelle part de marché pour les plus gros ?
Portefeuille	quelle part investie sur une seule ligne ?
Clients, fournisseurs	quelle part du chiffre d'affaires sur un seul client ?

Le dernier cas est un **risque de gestion** direct : une entreprise dont 60 % des ventes reposent sur un client est fragile, quelle que soit sa rentabilité.

Quand la concentration a un sens

La variable doit être **quantitative, positive et additive** : sa somme doit avoir une signification. La masse salariale a un sens ; la somme des âges ou des notes de satisfaction, non.

Dispersion contre concentration, en chiffres

Deux augmentations générales, sur nos 500 salaires

	Écart-type	Indice de Gini
Situation actuelle	2 598	0,195

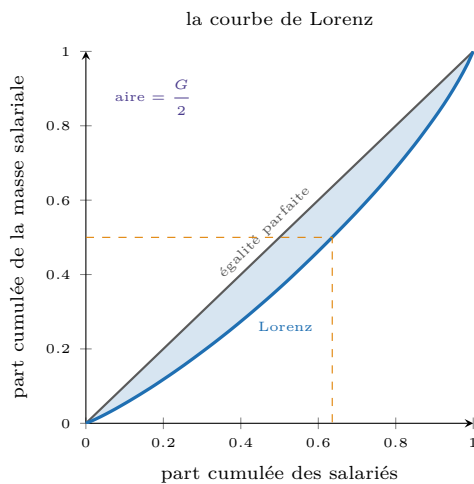
+1 000 dirhams pour tous	2 598	0,171
+10 % pour tous	2 858	0,195

Deux indicateurs, deux réponses

La prime uniforme laisse les **écarts** intacts mais réduit la **concentration** : chacun reçoit la même somme, qui pèse plus lourd pour les petits salaires. La hausse proportionnelle fait l'inverse. L'indice de Gini, défini plus loin, est insensible à un changement d'unité.

La courbe de Lorenz

La construction



Comment on la lit. L'abscisse classe les salariés du plus pauvre au plus riche. L'ordonnée donne la part de la masse salariale qu'ils se partagent.

Si tout le monde gagnait la même chose, la courbe serait la diagonale. Plus elle s'en écarte, plus la répartition est inégale.

Indice de Gini : le double de l'aire entre la courbe et la diagonale. $G = 0$: égalité parfaite. $G = 1$: une seule personne perçoit tout.

Sur nos données : $G = 0,195$.

La **médiane** vaut 7540 dirhams : les 64 % les moins payés se partagent la moitié de la masse salariale.

Comment on la trace

Les quatre étapes

1. Ordonner les individus par valeur **croissante**.
2. Calculer F_i , la part **cumulée des individus**.
3. Calculer Q_i , la part **cumulée de la masse** :

$$Q_i = \frac{\sum_{j \leq i} x_j}{\sum_j x_j}$$

4. Tracer les points $(F_i ; Q_i)$ dans le carré unité.

Trois propriétés

- La courbe part de $(0 ; 0)$ et arrive à $(1 ; 1)$.
- Elle est toujours **en dessous** de la diagonale, puisqu'on a ordonné par valeur croissante.

- Elle est **convexe** : chaque nouvel individu apporte plus que le précédent.

Lire la courbe

Sur nos salaires

Groupe	Part de la masse salariale
Les 10 % les moins payés	5,3 %
Les 50 % les moins payés	36,4 %
Les 20 % les mieux payés	31,4 %
Les 10 % les mieux payés	17,9 %

Comment on commente ces chiffres

« La moitié la moins payée du personnel se partage 36 % de la masse salariale ; les 10 % les mieux payés en captent 18 %. » Un responsable comprend immédiatement, alors qu'un écart-type de 2 598 dirhams ne lui dit rien. **La courbe de Lorenz est l'outil de communication le plus efficace de ce cours.**

Lorenz sur données groupées

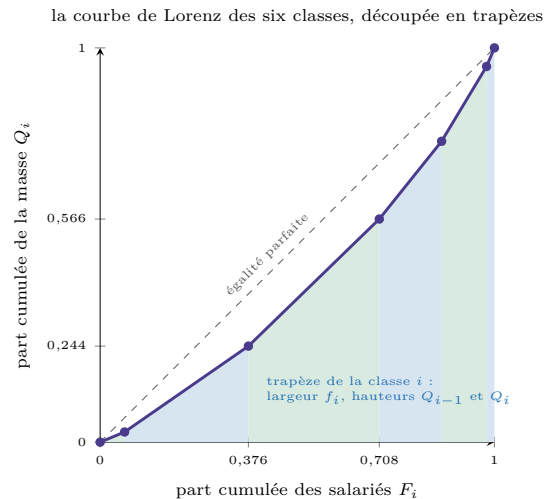
La masse de chaque classe, estimée par les centres (milliers de dirhams)

Classe	f_i	F_i	Masse $n_i c_i$	Part q_i	Q_i
[2 ; 4[	0,062	0,062	93	0,026	0,026
[4 ; 6[	0,314	0,376	785	0,218	0,244
[6 ; 8[	0,332	0,708	1 162	0,322	0,566
[8 ; 10[	0,158	0,866	711	0,197	0,763
[10 ; 14[	0,114	0,980	684	0,190	0,953
[14 ; 20[	0,020	1	170	0,047	1
Total	1		3 605	1	

Si la masse réelle est connue

Une publication donne souvent la masse de chaque classe : on l'utilise alors directement, sans passer par les centres.

La courbe des six classes



La lecture

Les 37,6 % de salariés des deux premières classes reçoivent 24,4 % de la masse.

L'indice de Gini

La définition

Indice de Gini

$$G = \frac{\text{aire entre la diagonale et la courbe}}{\text{aire du triangle sous la diagonale}} = 2 \times \text{aire entre la diagonale et la courbe}$$

G varie entre 0 et 1.

Les deux bornes

- $G = 0$: la courbe se confond avec la diagonale. **Égalité parfaite** — tout le monde reçoit la même chose.
- $G = 1$: la courbe longe les axes. **Concentration maximale** — un seul individu reçoit tout.

La formule de calcul

Sur données individuelles, ordonnées croissantes

$$G = \frac{2 \sum_{i=1}^n i x_i}{n \sum_{i=1}^n x_i} - \frac{n+1}{n}$$

Cinq valeurs : 2, 3, 4, 8, 13

$\sum i x_i = 1 \times 2 + 2 \times 3 + 3 \times 4 + 4 \times 8 + 5 \times 13 = 117$ et $\sum x_i = 30$:

$$G = \frac{2 \times 117}{5 \times 30} - \frac{6}{5} = 1,56 - 1,20 = 0,36$$

Sur nos 500 salaires, la même formule donne $G = 0,195$.

Le Gini par les trapèzes

Sur données groupées

L'aire sous la courbe est une somme de trapèzes, de largeur f_i et de hauteurs Q_{i-1} et Q_i .
D'où

$$G = 1 - \sum_{i=1}^k f_i (Q_{i-1} + Q_i)$$

D'où vient la formule

Un trapèze a pour aire $f_i \times (Q_{i-1} + Q_i)/2$. L'aire sous la courbe est la somme de ces trapèzes, et $G = 1 - 2 \times$ (aire sous la courbe) : le facteur 2 et le 1/2 se simplifient.

Le calcul sur les six classes

Les trapèzes des salaires

Classe	f_i	$Q_{i-1} + Q_i$	$f_i (Q_{i-1} + Q_i)$
$[2 ; 4[$	0,062	0,026	0,002
$[4 ; 6[$	0,314	0,269	0,085
$[6 ; 8[$	0,332	0,809	0,269
$[8 ; 10[$	0,158	1,329	0,210
$[10 ; 14[$	0,114	1,716	0,196
$[14 ; 20[$	0,020	1,953	0,039
Total			0,800

$G = 1 - 0,800 = 0,200$ — contre 0,195 sur les données exactes.

L'interprétation

Des repères indispensables

Un indice de Gini ne se lit **jamais seul** : il se compare.

Situation	Ordre de grandeur
Salaires au sein d'une entreprise	0,15 à 0,30
Revenus d'un pays très égalitaire	0,25 à 0,30
Revenus du Maroc	de l'ordre de 0,40
Revenus d'un pays très inégalitaire	0,50 à 0,60
Patrimoines (partout)	0,70 à 0,85

Ce que la comparaison enseigne

Une seule entreprise

Notre $G = 0,195$ est faible — mais il porte sur **une seule entreprise**, où l'éventail des postes est limité. Un Gini national additionne toutes les entreprises, tous les secteurs, et les revenus non salariaux. Les **patrimoines** sont toujours plus concentrés que les revenus : un stock accumule des décennies de flux inégaux.

Le Gini des salaires, par groupe

Groupe	Femmes	Hommes	Bac	Bac+2	Bac+5
Gini	0,213	0,181	0,181	0,155	0,165

Chaque groupe est **moins concentré que l'ensemble** pour le diplôme : une part de l'inégalité totale vient des écarts **entre** diplômes.

Ses limites

Deux distributions, un même Gini

Le Gini résume une courbe entière par un seul nombre. Deux courbes de Lorenz **qui se croisent** peuvent donner la même aire, donc le même indice, tout en décrivant des inégalités de nature différente : l'une concentrée sur le bas, l'autre sur le haut. **Toujours regarder la courbe avant de commenter l'indice.**

Les indices complémentaires

On publie souvent, à côté du Gini : le **rapport interdécile** D_9/D_1 , qui décrit les extrêmes ; la **part du dernier décile**, qui décrit le haut ; l'**indice de Theil**, décomposable entre groupes. Le Gini reste dominant parce que son interprétation graphique est immédiate.

La médiale

Une mesure complémentaire

Médiale

La **médiale** M_l est la valeur qui partage la **masse totale** en deux parties égales. À ne pas confondre avec la médiane, qui partage l'**effectif**.

Sur nos salaires

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

$$M_l = 7\,540 \text{ dirhams}$$

Les salariés gagnant moins de 7 540 dirhams représentent 63,6 % de l'effectif et se partagent exactement la moitié de la masse salariale.

La médiale sur données groupées

La même interpolation, sur la colonne des masses

On repère la classe où Q franchit 0,5 et l'on interpole :

$$M_l = e_{i-1} + a_i \times \frac{0,5 - Q_{i-1}}{Q_i - Q_{i-1}}$$

Le calcul

Q passe de 0,244 à 0,566 dans $[6\,000; 8\,000[$:

$$M_l = 6\,000 + 2\,000 \times \frac{0,500 - 0,244}{0,566 - 0,244} = 6\,000 + 1\,591 = 7\,591 \text{ dirhams}$$

Contre 7 540 sur les données exactes.

L'écart entre médiale et médiane

La propriété générale

$$M_l \geq Me$$

avec égalité si et seulement si la concentration est nulle.

Une mesure de concentration

Si tout le monde gagnait la même chose, partager l'effectif en deux reviendrait à partager la masse en deux. Plus les hauts salaires pèsent lourd, plus il faut monter haut pour atteindre la moitié de la masse :

$$\frac{M_l - Me}{Me} = \frac{770}{6\,770} = 11,4\%$$

La phrase qui résume

« Il faut réunir 64 % du personnel — et non 50 % — pour atteindre la moitié de la masse salariale. »

Bilan de la partie univariée

Les cinq familles

Ce que vous savez calculer

Famille	Question	Indicateurs
Tendance centrale	où ?	mode, médiane, moyennes
Quantiles	comment réparti ?	Q_1, Me, Q_3 , déciles
Dispersion	à quel point étalé ?	σ, IQ, CV
Forme	quelle allure ?	γ_1, γ_2
Concentration	comment partagé ?	Lorenz, Gini, médiale

Douze nombres pour cinq cents salariés

On est parti de 500 lignes illisibles. On arrive à douze nombres qui tiennent dans un coin de page et qui répondent à toutes les questions qu'on peut poser sur **une** variable : résumer, c'est perdre de l'information à bon escient.

La fiche complète de nos salaires

Douze nombres

Indicateur	Valeur	Indicateur	Valeur
Effectif	500	$Q_1 - Q_3$	5 318 — 8 330
Moyenne	7 158	D_9/D_1	2,41
Médiane	6 770	Asymétrie γ_1	1,16
Mode interpolé	6 188	Aplatissement γ_2	4,84
Écart-type	2 598	Gini	0,195
CV	0,363	Médiale	7 540

S'entraîner

Exercice 1 : la concentration des terres

1 000 exploitations agricoles

Superficie	Exploitations	Superficie totale (ha)
moins de 5 ha	600	1 200
de 5 à 20 ha	300	3 000
de 20 à 50 ha	80	2 400
50 ha et plus	20	3 400
Total	1 000	10 000

1. Calculer f_i , F_i , la part q_i des terres et Q_i .
2. Quelle part des terres détiennent les 2 % plus grandes exploitations ? Les 60 % plus petites ?
3. Calculer l'indice de Gini par la méthode des trapèzes et commenter.

Exercice 1 : solution

Le tableau de Lorenz

Hectares	f_i	F_i	q_i	Q_i	$f_i (Q_{i-1} + Q_i)$
< 5	0,60	0,60	0,12	0,12	0,0720
5 à 20	0,30	0,90	0,30	0,42	0,1620
20 à 50	0,08	0,98	0,24	0,66	0,0864
≥ 50	0,02	1	0,34	1	0,0332

Total	1	1	0,3536
-------	---	---	--------

Les réponses

2. Les 2 % plus grandes exploitations détiennent 34 % des terres ; les 60 % plus petites, 12 % seulement.
3. $G = 1 - 0,3536 = 0,65$: une concentration **très forte**, de l'ordre de celle des patrimoines — la terre est un patrimoine.

Exercice 2 : prime uniforme ou hausse proportionnelle

Quatre salariés

Une petite agence emploie quatre salariés, payés 3 000, 5 000, 8 000 et 14 000 dirhams.

1. Calculer l'indice de Gini par la formule $G = \frac{2 \sum i x_i}{n \sum x_i} - \frac{n+1}{n}$.
2. Chacun reçoit une prime de 1 000 dirhams. Recalculer G . L'écart-type a-t-il changé ?
3. Au lieu de la prime, chacun est augmenté de 10 %. Que deviennent G et l'écart-type ?

Exercice 2 : solution

Trois situations

1. $\sum i x_i = 3\,000 + 10\,000 + 24\,000 + 56\,000 = 93\,000$ et $\sum x_i = 30\,000$:

$$G = \frac{2 \times 93\,000}{4 \times 30\,000} - \frac{5}{4} = 1,55 - 1,25 = 0,30$$

2. Salaires 4 000, 6 000, 9 000, 15 000 : $\sum i x_i = 103\,000$, $\sum x_i = 34\,000$, d'où $G = 206\,000/136\,000 - 1,25 = 0,265$. L'écart-type, 4 153 dirhams, est **inchangé**.
3. Tout est multiplié par 1,1 : G reste à 0,30, l'écart-type passe à 4 569 dirhams.

La leçon

Pour **réduire les inégalités relatives**, la prime uniforme agit, la hausse proportionnelle non. Pour réduire les **écarts en dirhams**, aucune des deux.

Ce qu'il faut retenir

Le résumé du chapitre

Sept points

1. La **concentration** porte sur le partage de la **masse**, d'une variable **positive et additive**.
2. Une prime uniforme réduit la concentration sans changer l'écart-type.
3. La **courbe de Lorenz** trace la part cumulée de la masse contre la part cumulée des individus.
4. L'**indice de Gini** est le double de l'aire entre la courbe et la diagonale : 0 égalité, 1 concentration maximale.

5. Sur données groupées : $G = 1 - \sum f_i (Q_{i-1} + Q_i)$, par les trapèzes.
6. Un Gini ne se lit **jamais seul** : il se compare, et il ne remplace pas la courbe.
7. La **médiale** partage la masse ; elle dépasse la médiane, et l'écart mesure la concentration.

La suite

Les questions intéressantes en économie mettent en jeu **deux** variables : salaire **et** ancienneté, prix **et** quantité. C'est la partie IV : la statistique **bivariée**.