

Statistique descriptive

Chapitre 10 : La concentration



Jaouad Madkour

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

Une question nouvelle

La courbe de Lorenz

L'indice de Gini

La médiale

Bilan de la partie univariée

Ce qu'il faut retenir

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

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 ?

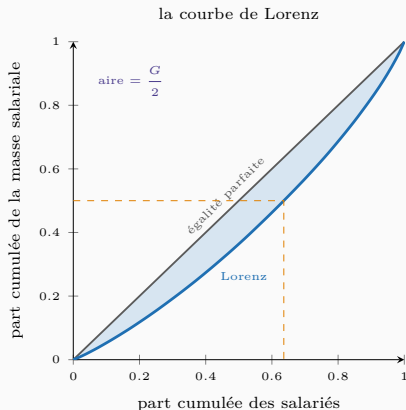
Le dernier cas est un **risque de gestion** direct : une entreprise dont 60 % des ventes reposent sur un client est fragile, quelle que soit sa rentabilité.

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, non. La somme des notes de satisfaction, non plus.

La courbe de Lorenz



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.

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** : la pente augmente, chaque nouvel individu apporte plus que le précédent.

La lecture 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 %. »

C'est une phrase qu'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 tout ce cours.

L'indice de Gini

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 des données individuelles ordonnées croissantes :

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

Sur nos salaires

$$G = 0,195$$

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

Notre $G = 0,195$ est faible — mais il porte sur les salariés d'**une seule entreprise**, où l'éventail des postes est nécessairement limité.

Un Gini national est bien plus élevé parce qu'il additionne toutes les entreprises, tous les secteurs, tous les statuts, et parce qu'il inclut les revenus non salariaux.

Et les **patrimoines** sont toujours plus concentrés que les revenus : le patrimoine est un stock qui accumule des décennies de flux inégaux.

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 de la distribution, l'autre sur le haut.

Toujours regarder la courbe avant de commenter l'indice — c'est le même avertissement qu'au chapitre 9 pour les moments, et qu'au chapitre 14 pour la corrélation.

Les indices complémentaires

C'est pourquoi 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 qu'il est le seul dont l'interprétation graphique soit immédiate.

La médiale

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 propriété générale

$$M_l \geq Me$$

avec égalité si et seulement si la concentration est nulle.

L'écart entre les deux est une mesure de concentration

Si tout le monde gagnait la même chose, partager l'effectif en deux reviendrait à partager la masse en deux : les deux valeurs coïncideraient.

Plus les hauts salaires pèsent lourd, plus il faut monter haut pour atteindre la moitié de la masse — donc plus M_l s'éloigne de Me .

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

La phrase qui résume

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

C'est le même fait que le Gini, dit autrement, et souvent plus parlant en réunion.

Bilan de la partie univariée

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

La fiche complète de nos salaires

Indicateur	Valeur
Effectif	500
Moyenne	7 158
Médiane	6 770
Mode (classe)	[6 000 ; 8 000[
Écart-type	2 598
CV	0,363
$Q_1 - Q_3$	5 318 - 8 330
D_9/D_1	2,41
Asymétrie γ_1	1,16
Aplatissement γ_2	4,84
Gini	0,195
Médiale	7 540

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.

C'est exactement le programme annoncé au chapitre 1 : résumer, c'est perdre de l'information à bon escient.

Ce qu'il faut retenir

Six points

1. La **concentration** porte sur le partage de la **masse**, pas sur l'écart entre individus.
2. Elle exige une variable quantitative **positive et additive**.
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. Un Gini ne se lit **jamais seul** : il se compare, et il ne remplace pas la courbe.
6. La **médiale** partage la masse ; elle est toujours supérieure ou égale à la médiane, et l'écart mesure la concentration.

La suite

Une variable est maintenant épuisée. Mais toutes les questions intéressantes en économie mettent en jeu **deux** variables : le salaire **et** l'ancienneté, le prix **et** la quantité, la publicité **et** les ventes.

C'est la partie IV : la statistique descriptive **bivariée**.