

STATISTIQUE DESCRIPTIVE

-Semestre 2-

Enseignante : Noura AIT ELMEKKI

Etablissement : ENCG MEKNES

Année Universitaire: 2019-2020

**CHAPITRE 1 : SÉRIE STATISTIQUE À
UNE SEULE VARIABLE
(Suite 1)**

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

- *Corrado Gini* est le statisticien derrière la notion de concentration. Ce statisticien italien a introduit cette notion notamment à propos des distributions de salaires et de revenus.
- L'objectif de la concentration est de mesurer les inégalités dans la répartition d'une variable à l'intérieur d'une population.
- La concentration s'applique à la description d'unités économiques selon la taille, par exemple des entreprises suivant le chiffre d'affaire, la production, le nombre de salarié... etc.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

- La concentration permet d'étudier la densité des données autour de la valeur centrale. La mesure de cette concentration concerne les variables statistiques *quantitatives continues* représentant une valeur *positive* cumulable.
- Les caractéristiques de concentration sont utiles pour mesurer la répartition de la *masse salariale*.
- La répartition de la masse salariale se situe entre les deux cas extrêmes suivants:
 - *Concentration nulle*: la répartition des salaires est parfaitement équitable; un certain pourcentage de salariés reçoit le même pourcentage de la masse salariale.
 - *Concentration maximale*: un seul salarié reçoit toute la masse salariale (et les autres ne reçoivent rien).

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Les caractéristiques de concentration, se présentent comme suit:

- **Écart entre médiale et médiane**
- **La courbe de concentration appelée Courbe de Lorentz**
- **L'indice de Gini.**

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

- Néanmoins, pour procéder au calcul de ces caractéristiques il convient d'abord de déterminer:

→ **Les valeurs globales (les masses)**

→ **La médiale**

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Les masses

- Soit une série statistique où la variable statistique X étudiée est continue et dont les valeurs sont positives et regroupées en k classes, d'effectifs respectifs n_1, n_2, \dots, n_k , et de centres respectifs c_1, c_2, \dots, c_k . On appelle:
- S_i , la *masse* associée à la classe $[e_i; e_{i+1}[$, d'effectif n_i , tel que:

$$S_i = n_i c_i$$

- S , la *masse globale* de la variable X , tel que:

$$S = \sum_{i=1}^k S_i = \sum_{i=1}^k n_i c_i$$

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Les masses

- g_i , la *masse relative* (appelée également proportion de la masse globale) associée à la classe $[e_i; e_{i+1}[$, tel que:

$$g_i = \frac{S_i}{S} = \frac{n_i c_i}{\sum_{i=1}^k n_i c_i}$$

- G_i , la *masse relative cumulée croissante* associée à la classe $[e_i; e_{i+1}[$, tel que:

$$G_i = \sum_{i=1}^k g_i$$

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Les masses

Exemple: le tableau suivant présente la distribution de salaire de 50 employés d'une entreprise:

Salaire (en DH)	ni	ci	fi (%)	Fi (FCC)
[600;1200[15	900	30	30
[1200;1800[25	1500	50	80
[1800;2100[10	1950	20	100
Total	50	-	100	-

Calculer la masse salariale globale, ainsi que les masse relatives et les masses relatives cumulées croissantes.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Les masses

Exemple: correction

Salaire (en DH)	ni	ci	fi %	Fi % (FCC)	si (nici)	gi % (si/S)	Gi %	
[600; 1200[15	900	30	30	13500 ¹	19.15 ²	19.15	<p>Masse salariale relative cumulée croissante</p>
[1200;1800[25	1500	50	80	37500	53.19	72.34	
[1800;2100[10	1950	20	100	19500	27.66	100	
Total	50	-	100	-	70500	100	-	

(1) $15 \cdot 900 = 13500$

(2) $13500 / 70500 \cdot 100 = 19.15\%$

Masse
salariale
globale (S)

Masse
salariale
relative

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

La médiale

- La médiale d'une série statistique est la valeur du caractère X qui partage en deux parties égales la masse totale du caractère. La médiale s'exprime dans la même unité que la variable étudiée:

$$M_l = G(M_l) = 0.5 = 50\%$$

- Comme la médiane, la médiale est calculée par interpolation linéaire à partir d'une classe médiale. La classe médiale est la classe dont la masse salariale relative cumulée croissante est supérieure ou égale à 50% ($G(e_i) < 0.5$ et $G(e_{i+1}) \geq 0.5$).

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

La médiale

- *Attention!* Ne pas confondre médiane et médiale.
- La médiale est la médiane de la série masse associée.
- La médiale de la série (x_i, n_i) est la médiane de la série (x_i, g_i) .
- La médiale est calculée par la formule suivante:

$$M_l = e_i + a_i \frac{50 - G(e_i)}{G(e_{i+1}) - G(e_i)}$$

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

La médiale

- *Exemple précédent: calculer la médiale*

Salaire (en DH)	ni	ci	fi %	Fi % (FCC)	si (nici)	gi % (si/S)	Gi %
[600; 1200[15	900	30	30	13500	19.15	19.15
[1200;1800[25	1500	50	80	37500	53.19	72.34
[1800;2100[10	1950	20	100	19500	27.66	100
Total	50	-	100	-	70500	100	-

Classe médiale: [1200;1800[

$$M_l = 1200 + 600 \frac{50 - 19.15}{72.34 - 19.15}$$

Les salariés recevant moins de 1548 DH représentent 50% de la masse salariale.

$$M_l \cong 1548 \text{ DH}$$

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Écart médiale - médiane

- La différence entre la médiale et la médiane, $M_l - M_e$, est une mesure de la concentration d'une distribution.

ce rapport est compris entre 0 et 1

$$\Delta = \frac{M_l - M_e}{E}$$

Médiale ——— Médiane
Étendue

- Concentration forte**: lorsque cette différence est grande par rapport à l'étendue.
- Concentration faible**: lorsque cette différence est petite.
- Concentration nulle**: la médiane est égale à la médiale ($M_l - M_e = 0$). Dans ce cas la distribution est *égalitaire*: l'ensemble des x_i sont égaux.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Écart médiale-médiane

- Exemple précédent: calculer l'écart médiale-médiane.

Salaire (en DH)	ni	ci	fi %	Fi % (FCC)	si (nici)	gi % (si/S)	Gi %
[600; 1200[15	900	30	30	13500	19.15	19.15
[1200;1800[25	1500	50	80	37500	53.19	72.34
[1800;2100[10	1950	20	100	19500	27.66	100
Total	50	-	100	-	70500	100	-

Médiale: $M_l = 1548$

$$\Delta = (1548 - 1440) / (2100 - 600)$$

Médiane: $M_e = 1440$

$$\Delta = 7.2\%$$

L'écart médiale-médiane représente 7.2% de l'étendue. Ce qui montre que la concentration des salaires est faible.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

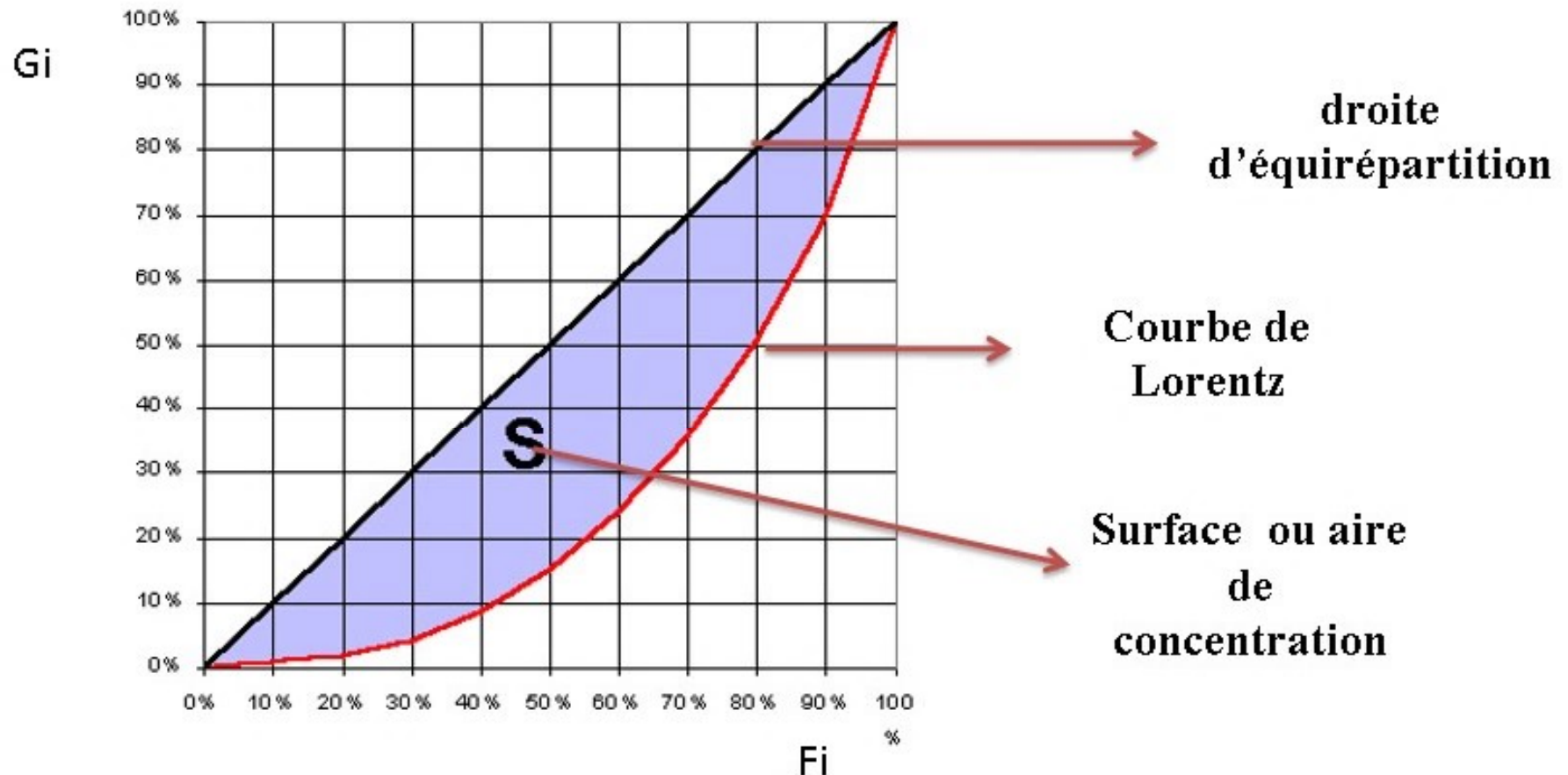
Courbe de Lorentz

- La courbe de concentration, appelée courbe de Lorentz, est obtenue à partir des calculs des Fréquences cumulées croissantes (F_i) et des masses cumulées croissantes (G_i).
- Pour tracer cette courbe on pose:
 - *En abscisse*, les fréquences cumulées croissantes de la série (x_i, n_i) : **F_i** .
 - *En ordonnée*, les fréquences cumulées croissantes de la série (x_i, g_i) : **G_i** .
- On trace par la suite la *droite d'équirépartition* ou droite d'égalité parfaite. Cette droite présente une situation d'égalité où la répartition est parfaitement équitable.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Courbe de Lorentz

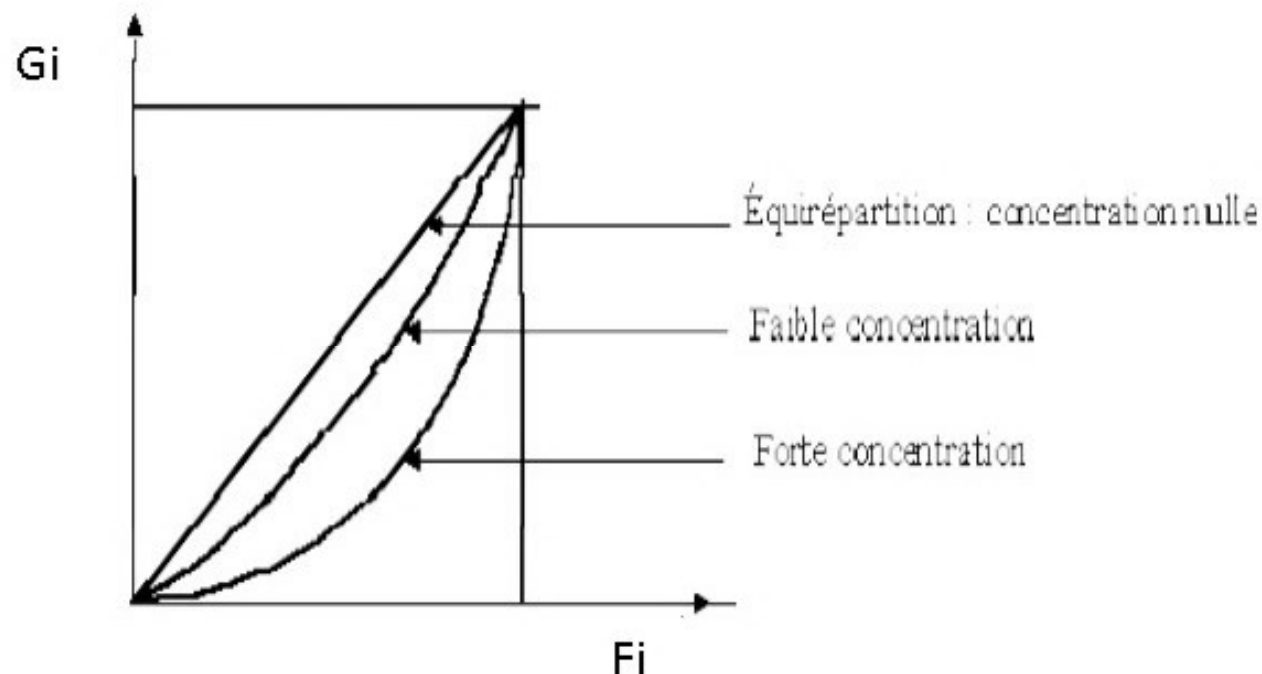


CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Courbe de Lorenz

- L'allure de la courbe nous permet d'avoir une idée sur la concentration.



→ Pour une *concentration faible*, la surface comprise entre la courbe de Lorenz et la droite d'équirépartition, est petite.

→ Pour *une concentration forte*, cette surface est grande.

→ Pour *une concentration nulle*, la courbe de Lorenz et la droite d'équirépartition sont confondues.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Courbe de Lorenz

- *Exemple précédent : tracer la courbe de Lorenz.*



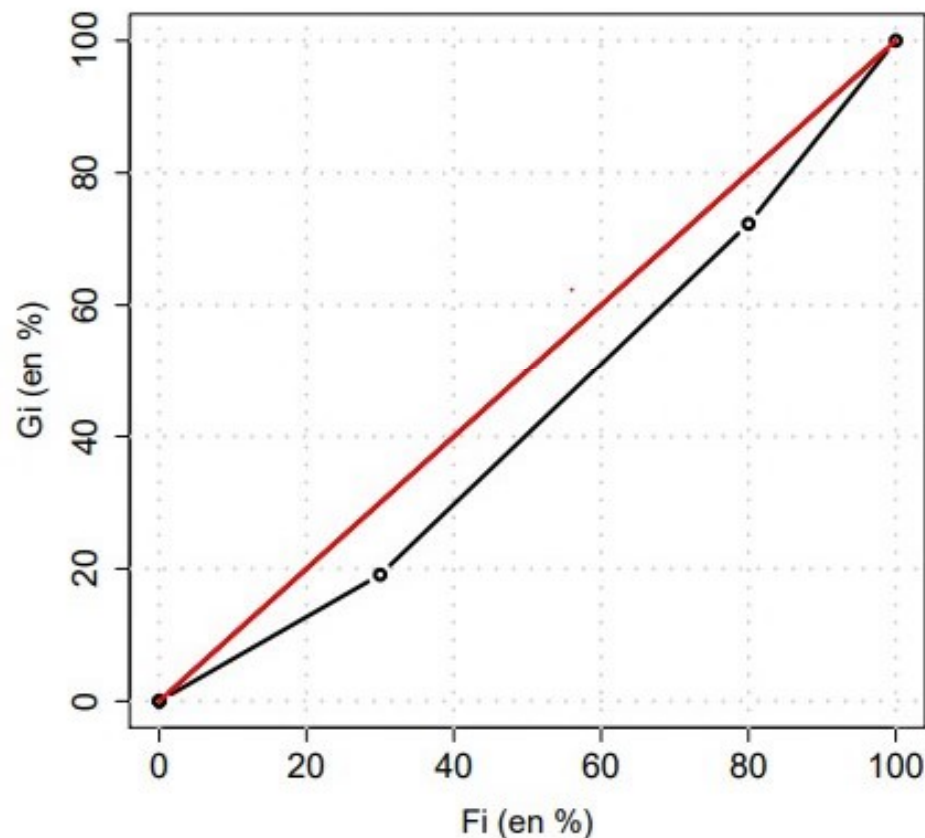
Salaire (en DH)	ni	ci	fi %	Fi % (FCC)	si (nici)	gi % (si/S)	Gi %
[600; 1200[15	900	30	30	13500	19.15	19.15
[1200;1800[25	1500	50	80	37500	53.19	72.34
[1800;2100[10	1950	20	100	19500	27.66	100
Total	50	-	100	-	70500	100	-

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Courbe de Lorenz

- *Exemple précédent : tracer la courbe de Lorenz.*



— Droite d'équirépartition
— Courbe de Lorenz

La surface de concentration n'est pas grande donc la concentration des salaires est plutôt faible. La répartition des salaires est par conséquent plus au moins équitable.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Indice de Gini

- L'indice de Gini est une mesure du degré d'inégalité.
- Il correspond au double de la surface comprise entre la courbe de concentration et la droite d'équirépartition.
- Algébriquement, l'indice de Gini est calculé à travers la formule suivante:

$$I_G = 1 - \sum_{i=1}^k f_i (G_i + G_{i-1})$$

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Indice de Gini

- L'indice de Gini est compris entre 0 et 1 (0 et 100 s'il est exprimé en %):
 - $I_G=0$: Égalité parfaite; la courbe de Lorenz et la droite d'équirépartition sont confondues.
 - $I_G=1$: Inégalité maximale; une seule personne détient tout le revenu.
 - I_G proche de 0: faible concentration.
 - I_G proche de 1: forte concentration.
- Plus les inégalités sont importantes, plus la courbe de Lorenz s'éloigne de la droite, plus l'indice de Gini est élevé.

CHAPITRE 1 : SÉRIE STATISTIQUE À UNE SEULE VARIABLE

4. Caractéristiques de concentration

Indice de Gini

- Exemple précédent : calculer l'indice de Gini.



Salaire (en DH)	f_i %	F_i % (FCC)	s_i (nici)	g_i % (s_i/S)	G_i %	(G_i+G_{i-1}) %	$f_i(G_i+G_{i-1}) 10^{-4}$
[600; 1200[30	30	13500	19.15	19.15	19.15 ⁽¹⁾	574.5 ⁽⁴⁾
[1200;1800[50	80	37500	53.19	72.34	91.49 ⁽²⁾	4574.5 ⁽⁵⁾
[1800;2100[20	100	19500	27.66	100	172.34 ⁽³⁾	3446.8 ⁽⁶⁾
Total	100	-	70500	100	-	-	8595.8

(1): $19.15+0$

(2): $72.34+19.15$

(3): $100+72.34$

(4): $30*19.15$

(5): $50*91.49$

(6): $20*172.34$

$I_G = 1 - 0.86 = 0.14.$

$I_G = 14\%$. La concentration est donc faible, et par conséquent la répartition des salaires est plus au moins équitable.