Université Abdelmalek Essaadi
ENCG - TANGER

Semestre 2

chap6 : Les caractéristiques de concentration

Prof: Abdelali ZBAKH <a.zbakh@uae.ac.ma>

Les caractères de concentration

- La mesure de la concentration revient à celle de la conséquence de dispersion. Très importante en économie (concentration des salaires, des revenus, de la taille des entreprises...) elle concerne des variables continues ne pouvant prendre que des valeurs positives.
- L'objectif est de mesurer les inégalités dans la répartition d'une variable à l'intérieur d'une population.
- Il existe deux méthodes de détermination de concentration:
  - ✓ par calcul;
  - ✓ par les graphes.
- Caractéristiques de concentration :
  - ✓ Médiale
  - ✓ Courbe de Lorentz
  - ✓ Indice de Gini

Détermination analytique de la concentration: Valeurs globales

Considérons une série statistique comportant n observations dans un tableau statistique $(x_i; n_i)$, présentant k modalités, on appelle :

- ✓ **Masse** associée à la modalité x_i d'effectif n_i la quantité définie par:

$$\text{masse} = n_i x_i$$

- ✓ **Masse totale**:

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

- ✓ **Masse relative** associée à la modalité x_i la quantité définie par:

$$q_i = \frac{n_i x_i}{S}$$

Détermination analytique de la concentration: Valeurs globales

Exemple: **Données discrètes groupées**.

Pour la distribution suivante d'une population de 16 familles selon le nombre d'enfants

Nombre d'enfants par famille x_i	Nombre de famille n_i	masse $n_i x_i$	Fréquence des masses du caractère (qi en %)	Fréquence cumulée croissante des masses du caractère (qi cc en %)
0	1	0	0.00	0.00
1	3	3	9.68	9.68
2	8	16	51.61	61.29
3	4	12	38.71	100.00
Total	16	$\boxed{31}$	100	-

Masse totale

Statistique Descriptive

Détermination analytique de la concentration: Valeurs globales

$$\text{masse} = n_i c_i$$

Exemple: **Données groupées en classes**

Masse totale: $S=\sum_{i=1}^{k} n_i C_i$

Taille (en m)	n_i	Ci	masse niCi	Fréquence des masses du caractère (qi en %)	Fréquence cumulée croissante des masses du caractère (qi cc en %)
[1,50 ; 1,60[	6	1.55	9.3	22.96	22.96
[1,60 ; 1,70[	7	1.65	11.55	28.52	51.48
[1,70 ; 1,80[	8	1.75	14	34.57	86.05
[1,80 ; 1,90[	2	1.85	3.7	9.14	95.19
[1,90 ; 2[	1	1.95	1.95	4.81	100.00
Total	24	--	$\boxed{40.5}$	100.00	--

Masse totale

Statistique Descriptive

Détermination analytique de la concentration: La Médiale

- La Médiale est la valeur du caractère qui partage en deux parties égales **la masse totale** du caractère ($n_i x_i$). La Médiale est notée Ml

- La **médiale** d'une série ($x_i;n_i$) est la **médiane** de la série ($\mathbf{x_i}$,$\boldsymbol{n_i x_i}$).

Remarques :

- La masse relative $\mathbf{q_i}$ de la série (x_i,$n_i x_i$) joue le rôle de la fréquence $\mathbf{f_i}$ dans la série (x_i,n_i)
- Pour calculer la Médiale, on applique la même formule que la **médiane** mais appliquée aux données $n_i x_i$ ou q_i.

Statistique Descriptive

Détermination analytique de la concentration: La Médiale

Exemple: **Données discrètes groupées**.

Nombre d'enfants par famille x_i	Nombre de famille n_i	masse $n_i x_i$	masse CC
0	1	0	0
1	3	3	3
2	8	16	$\boxed{19}$
3	4	12	31
Total	16	31	

S=31 est impair, La **médiale** est la valeur qui occupe le $\frac{31+1}{2}$=16 rang de la série ordonnée (xi,xi ni).

La médiale est alors égale à 2.

Statistique Descriptive

Exemple: **Données groupées en classes**

Taille (en m)	n_i	Ci	masse niCi	Fréquence des masses du caractère (qi en %)	Fréquence cumulée croissante des masses du caractère (qi cc en %)
[1,50 ; 1,60[	6	1.55	9.3	22.96	22.96
[1,60 ; 1,70[	7	1.65	11.55	28.52	51.48
[1,70 ; 1,80[	8	1.75	14	34.57	86.05
[1,80 ; 1,90[	2	1.85	3.7	9.14	95.19
[1,90 ; 2[	1	1.95	1.95	4.81	100.00
Total	24	--	40.5	100.00	--

- La classe médiale [a,b[est celle dont la qi cc (b) est ≥ 50% (ou masse cumulée Mcc (b) est ≥S/2) et dont la classe précédente a une qi cc (a) < 50 % (ou masse cumulée Mcc (a) est < S/2).
- La médiale est calculée par la formule:

$$Ml = a + \frac{(50 - qicc(a))}{qicc(b) - qicc(a)} \times (b - a)$$

$$Ml = a + \frac{(S/2 - Mcc(a))}{Mcc(b) - Mcc(a)} \times (b - a)$$

- la classe médiale est [1,60 ; 1,70[, Ml appartient à [1,60 ;1,70[

$$Ml = 1{,}60 + \frac{(50 - 22.96)}{51.48 - 22.96} \times (1{,}70 - 1{,}60) = 1{,}6948 \text{ m}$$

Interprétation : La moitié de la taille totale des 24 étudiants est réalisée par les étudiants dont la taille est inférieure à 1,6948 m.

Application 1: (A faire)

- Déterminer la médiale par calcul
- Déterminer la médiale graphiquement, en utilisant la courbe des fréquences cumulées des masses qi cc .

Taille (en m)	n_i	Ci	masse niCi	Fréquence des masses du caractère (qi en %)	Fréquence cumulée croissante des masses du caractère (qi cc en %)
[1,50 ; 1,60[	6	1.55	9.3	22.96	22.96
[1,60 ; 1,70[	7	1.65	11.55	28.52	51.48
[1,70 ; 1,80[	8	1.75	14	34.57	86.05
[1,80 ; 1,90[	2	1.85	3.7	9.14	95.19
[1,90 ; 2[	1	1.95	1.95	4.81	100.00
Total	24	--	40.5	100.00	--

Statistique Descriptive

Détermination analytique de la concentration:

Une mesure de la concentration :

$$C = \frac{\Delta M}{l'étendue}$$

Avec

- $\Delta M = Ml - Me$: *est* l'écart médiale-médiane

- **Si** $\Delta\ M = 0$ alors la concentration est nulle.
- **Si** ΔM est grande par rapport à l'étendue $\Leftrightarrow$ $C > 1$, alors la concentration est forte ;
- **Si** ΔM est petite par rapport à l'étendue $\Leftrightarrow$ $C < 1$, alors la concentration est faible.

Détermination analytique de la concentration:

Exemple:

salaires	n_i	$x_i=C_i$	$n_{i\,cc}$	masse $n_i\,x_i$	masse cc
[10 – 15[	9	12,5	9	112,5	112,5
[15 – 20[	25	17,5	34	437,5	550
[20 – 25[	32	22,5	66	720	1270
[25 – 30[	16	27,5	**82**	440	1710
TOTAL	**82**			**1710**	

Colonne *ni cc:*

k =82/2 = 41 donc **la classe médiane** est **[20 ; 25[**
On applique la formule de calcul de la médiane :

$$Me= 20 + \frac{(41 - 34)}{66 - 34} \times (25 - 20) = 21,09$$

Colonne *masse cc*

k =1710/2 = 855 donc **la classe médiale** est **[20 ; 25[**
On applique la formule de calcul de la médiale :

$$Ml= 20 + \frac{(855 - 550)}{1270 - 550} \times (25 - 20) = 22,12$$

La mesure de la concentration: $C=\frac{\Delta M}{Etendue} = \frac{22,12 - 21,09}{30-10} = 0.0515 < 1$, donc la concentration est faible

Détermination géométrique de la concentration : la courbe de concentration

Détermination graphique de la concentration:

- Il existe un moyen visuel de déterminer la concentration sans passer par la comparaison de (médiane, médiale).
- Il suffit de confronter les deux fonctions cumulatives sur un graphique, appelé : (courbe de Lorentz ou courbe de Gini ou courbe de Gini-Lorentz)

<u>Méthode de construction</u> :

- Pour tracer la courbe de Gini-Lorenz on met :

✓ Sur l'axe des abscisses la fréquence cumulée ($\boldsymbol{f_{i\,cc}}$) de la série $(x_i; n_i)$;

✓ Sur l'axe des ordonnées la fréquence cumulée ($\boldsymbol{q_{i\,cc}}$) de la série $(x_i; n_i \times x_i)$;

✓ On trace la droite passant par l'origine d'équation $q_{icc} = f_{icc}$.

Statistique Descriptive

La courbe de Gini-Lorenz

- La série suivante indique la répartition des 22 régions françaises selon le nombre de lits dont elles disposent en maisons de retraite au 1 Janvier 2005 :

Nombre de lits	n_i	$x_i = C_i$	f_i	f_{iCC}	masse $n_i x_i$	q_i (%)	q_{iCC}
[0 ; 12 250[	4	6 125	18,18%	18,18%	24 500	5,41%	5,41%
[12 250 ; 24 500[	12	18 375	54,55%	72,73%	220 500	48,65%	54,05%
[24 500 ; 36 750[	4	30 625	18,18%	90,91%	122 500	27,03%	81,08%
[36 750 ; 49 000[	2	42 875	9,09%	100,00%	85 750	18,92%	100,00%
Total	22		100,00%		453 250	100,00%	

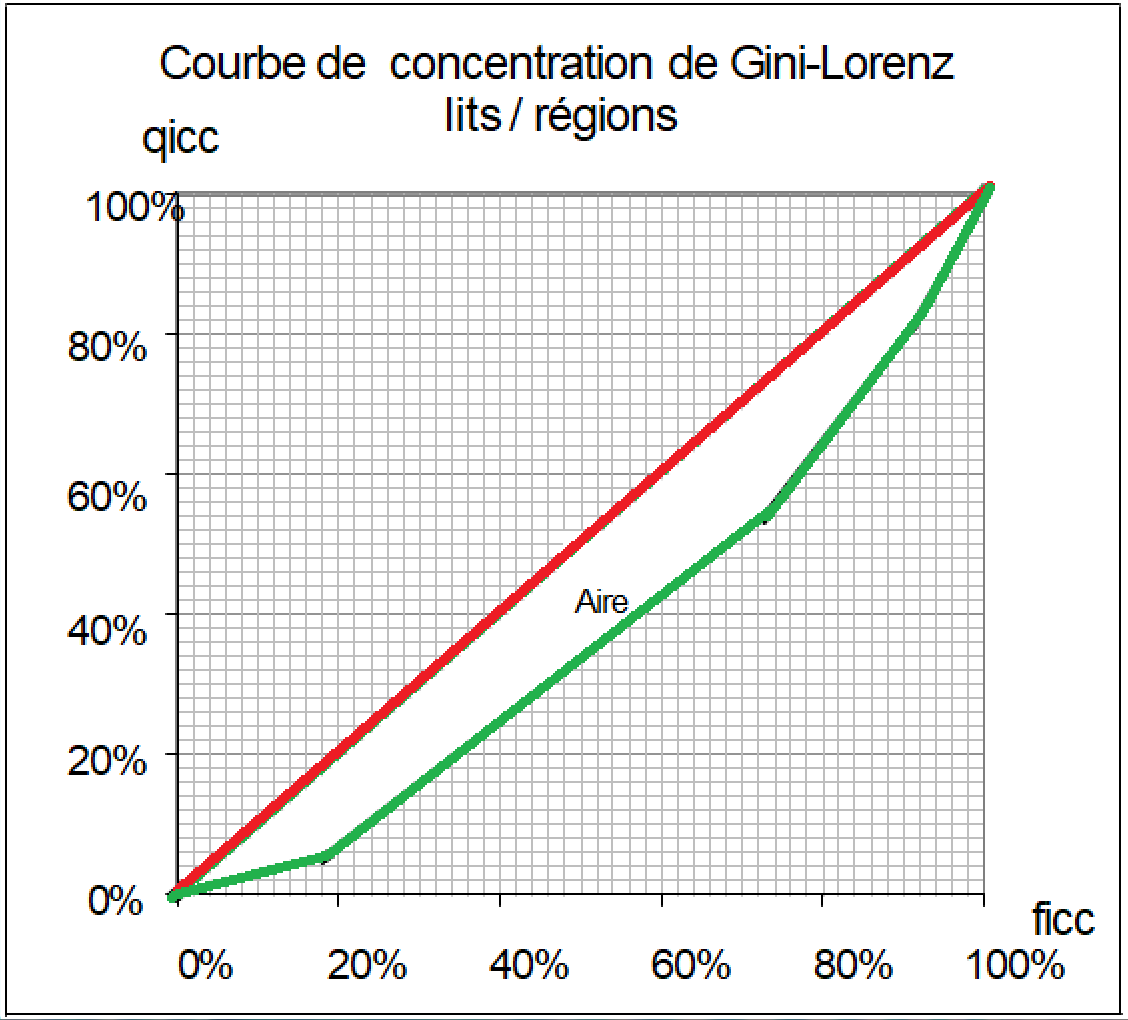
Calcul de la médiale : la classe médiale c'est **[12 250 ; 24 500[** On trouve **Ml= 23 479,17 lits**

Ce qui signifie que 50% des lits disponibles en maisons de retraite françaises proviennent de régions qui ont moins de 23 479,17 lits.

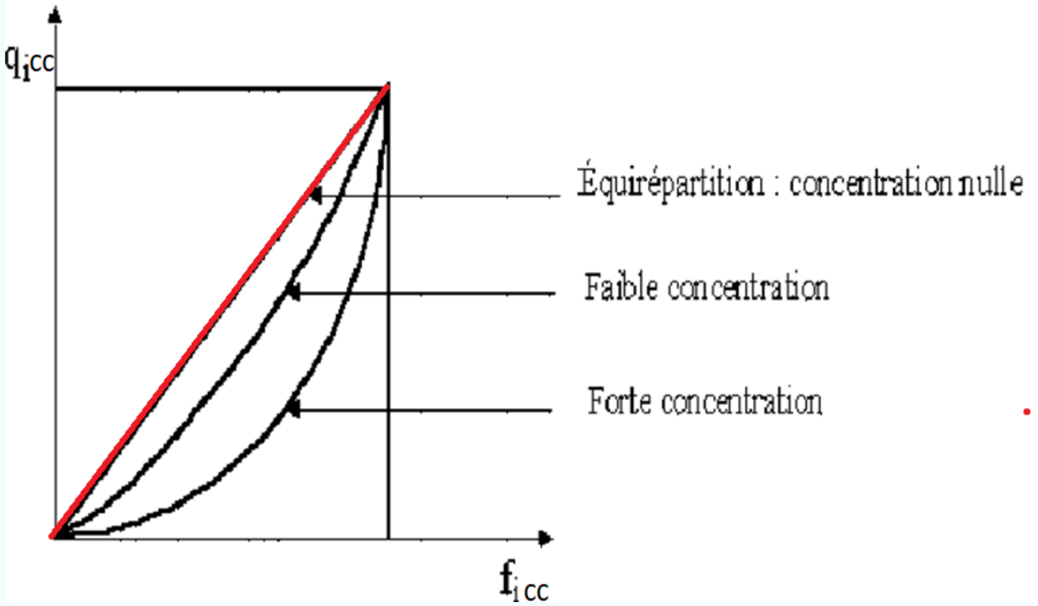
Pour construire la courbe de Gini-Lorenz on place les points de cordonnées :

(18,18 ; 5,41) - (72,73 ; 54,05) - (90,91 ; 81,08) - (100 ; 100)

La courbe de Gini-Lorenz



- La courbe de Lorenz illustre la répartition de la richesse dans une société.
- Plus cette courbe est éloignée de la bissectrice, plus les inégalités sont fortes.
- Le coefficient de Gini mesure les inégalités au sein d'une société.

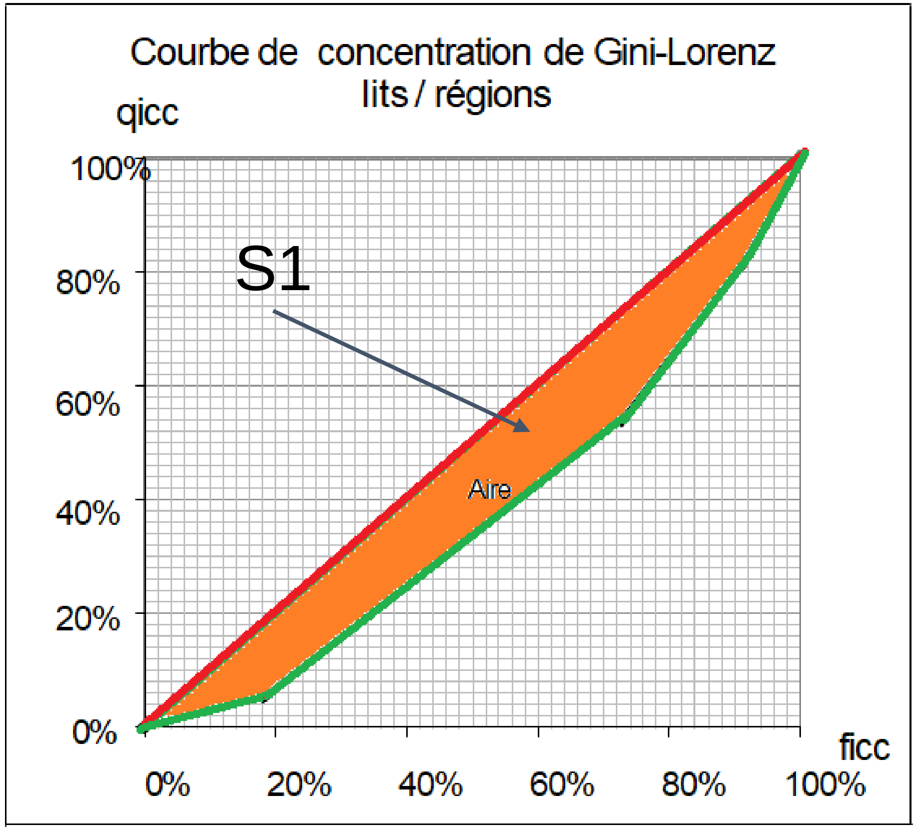


La courbe de Gini-Lorenz: L'indice de Gini

- L'indice de GINI est une mesure de l'indice de concentration à partir de la courbe de Gini-Lorenz. Il est donné par la formule :

$$G = 2 \times S1$$

- où $S1$ est l'**aire** comprise entre la courbe de concentration est la diagonale.



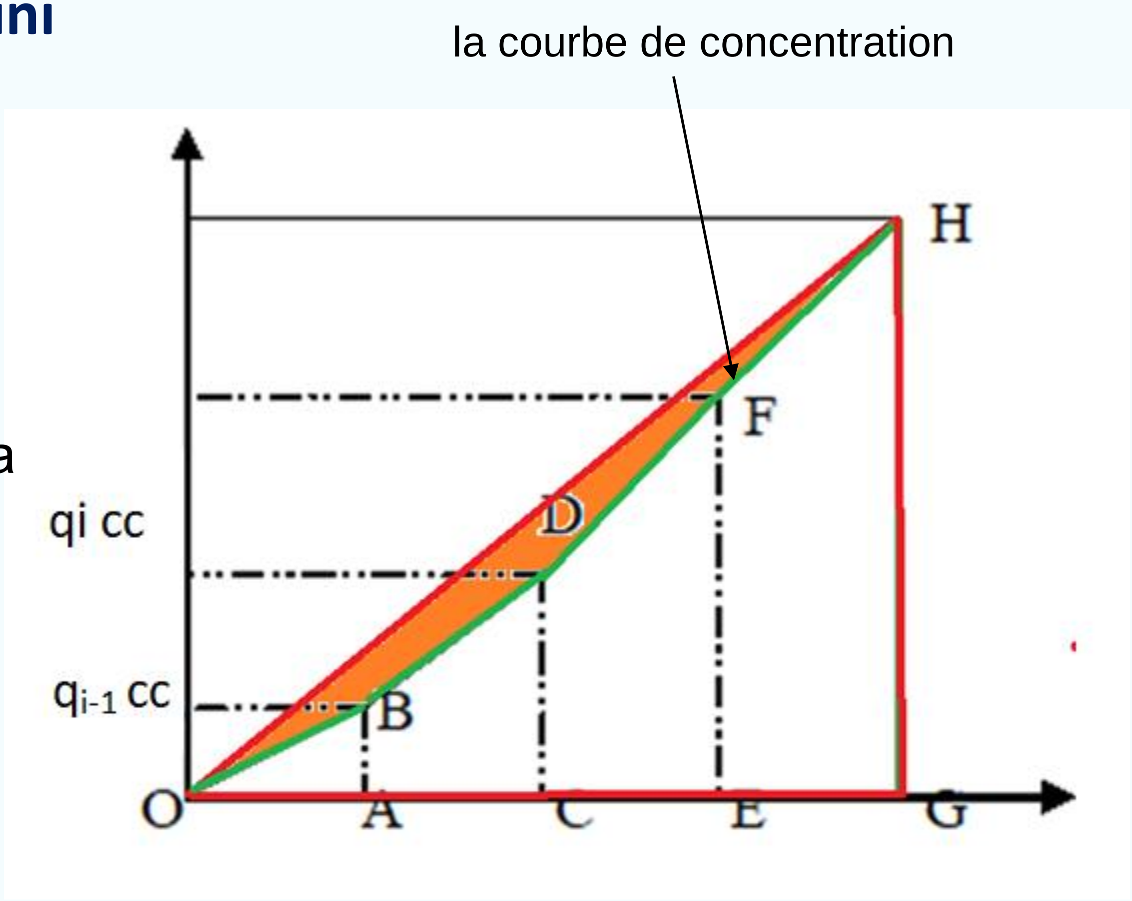
Propriétés de l'indice de Gini

- On a toujours $\mathbf{0} < \mathbf{G} < \mathbf{1}$;
- G proche de 1 $\Longrightarrow$ forte concentration(plus la société est inégalitaire) ;
- G proche de 0 $\Longrightarrow$ faible concentration.

La courbe de Gini-Lorenz: L'indice de Gini

Calcul de l'indice de Gini : **Méthode des Trapèzes**

- l'aire comprise entre la courbe de concentration et l'axe des abscisses peut être divisée en autant de trapèzes qu'il y a de modalités.
- Sur ce graphique peuvent être mis en évidence 4 trapèzes successivement (OAB), (ACDB), (CEFD), (EGHF).



Statistique Descriptive

La courbe de Gini-Lorenz: L'indice de Gini

Calcul de l'indice de Gini : **Méthode des Trapèzes**

On note :

- S1=l'Aire de concentration.
- S2= Aire du triangle OGH $=\frac{1 \, x \, 1}{2}$
- S3=∑des surfaces des trapèzes

S2= S3+ S1

S1= S2- S3

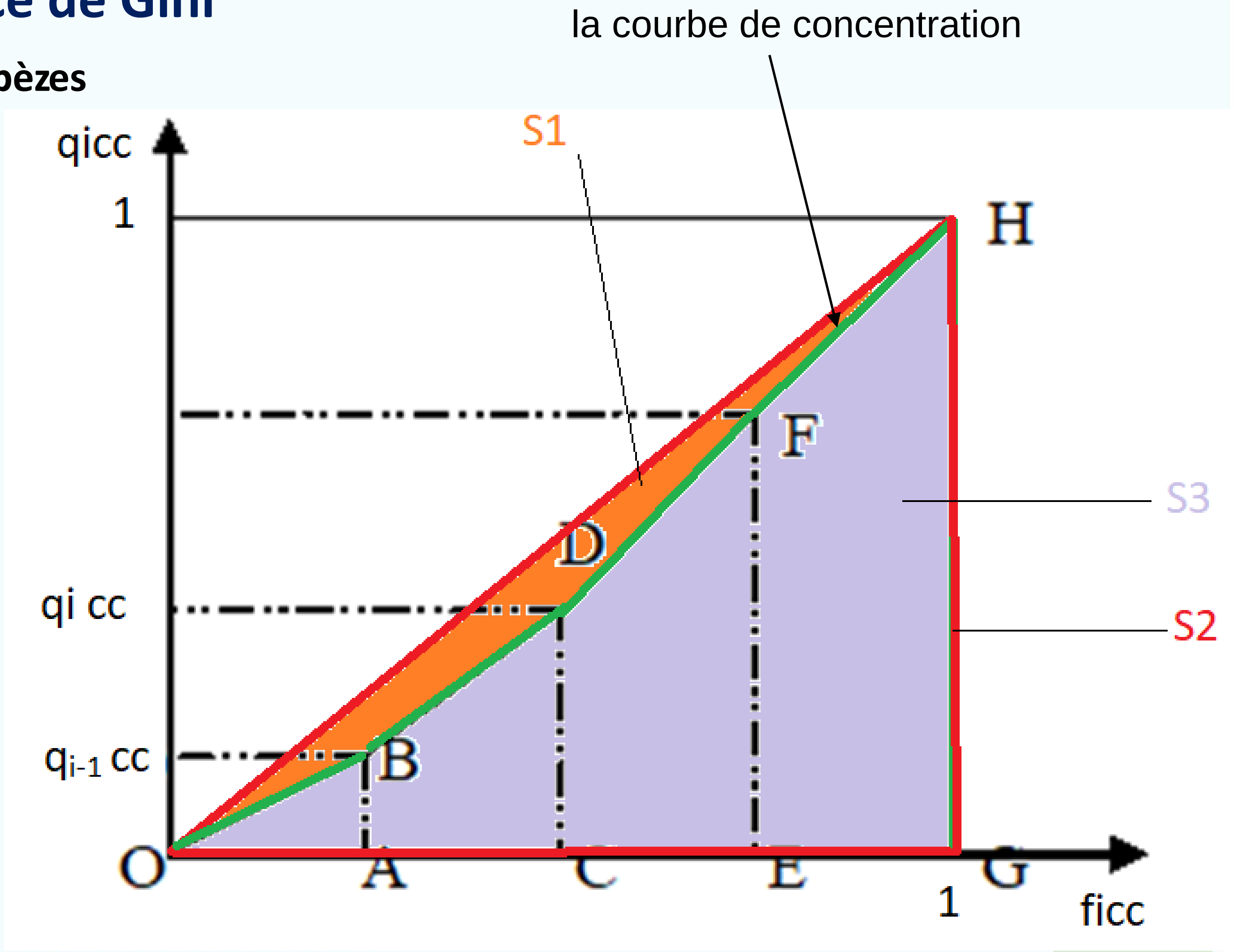
S1= $\frac{1}{2}$**- S3**

L'indice de GINI;

G = 2 × S1

G = 2 ×($\frac{1}{2}$**- S3)**

G =1-2 x S3



La courbe de Gini-Lorenz: L'indice de Gini

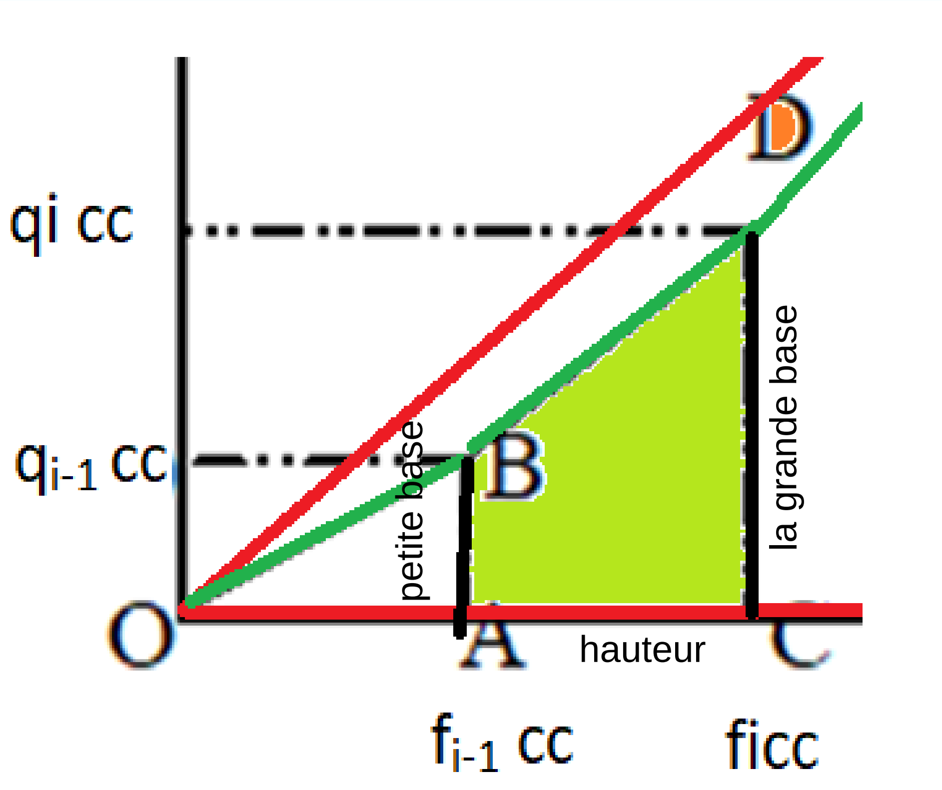
Calcul de l'indice de Gini : **Méthode des Trapèzes**

$$\text{Aire trapèze} = \frac{(\text{la grande base} + \text{la petite base}) \times \text{hauteur}}{2}$$

$$= \frac{(q_i cc + q_{i-1} cc) \times (f_i cc - f_{i-1} cc)}{2}$$

$$= \frac{(q_i cc + q_{i-1} cc) \times f_i}{2}$$

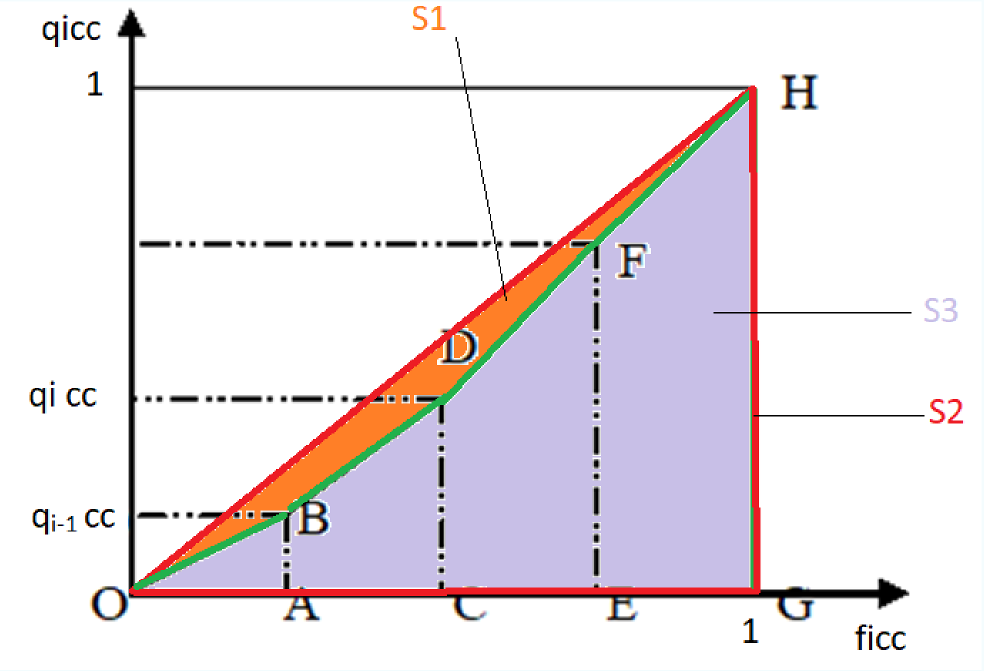
avec $f_i cc = f_i + f_{i-1} cc$



Statistique Descriptive

La courbe de Gini-Lorenz: L'indice de Gini

Calcul de l'indice de Gini : **Méthode des Trapèzes**



Aire de tous les trapèzes :

$$S3 = \sum \frac{(qicc + qi_{-1}cc) \times fi}{2}$$

L'indice de GINI:

$$G = 1 - 2 \times S3 = 1 - 2 \times \sum fi \times (q_{i-1}cc + q_i cc)/2$$

$$= 1 - \sum fi \times (q_{i-1}cc + q_i cc)$$

La courbe de Gini-Lorenz: L'indice de Gini

- Exemple :calculer l'indice de Gini

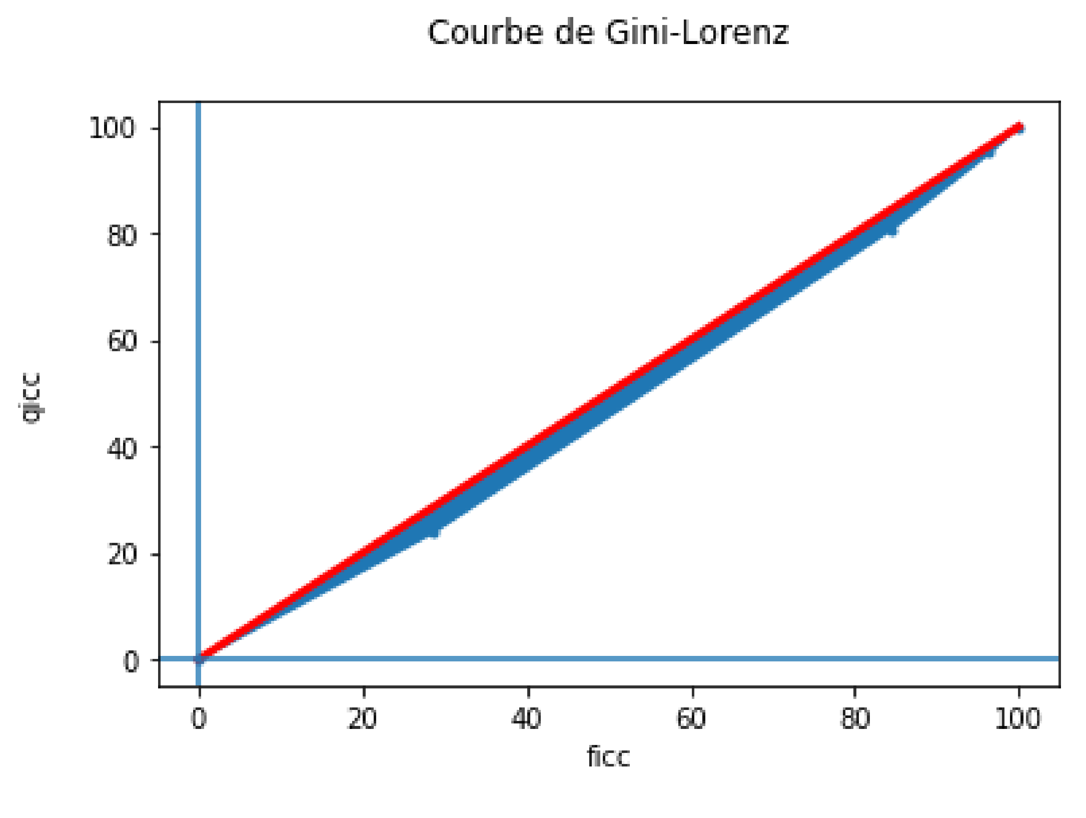
xi	ni	ci	ai	ni x ci	Nicc	fi	Ficc	qi	qicc	$s_i=q_{i-1}cc + q_i\,cc$	si x fi
[1000;1250[	24	1125	250	27000	24	0.285714286	0.2857	0.237885463	0.237885463	0.237885463	0.06796728
[1250;1500[	47	1375	250	64625	47	0.55952381	0.8452	0.56938326	0.807268722	1.045154185	0.58478865
[1500;1750[	10	1625	250	16250	10	0.119047619	0.9643	0.143171806	0.950440529	1.757709251	0.2092511
[1750;2000[	3	1875	250	5625	3	0.035714286	1	0.049559471	1	1.950440529	0.06965859
somme	84			113500	84	1		1			0.93166562

L'indicede Gini est = **1** - 0.93166562 = **0.068334** → **La concentration est faible**

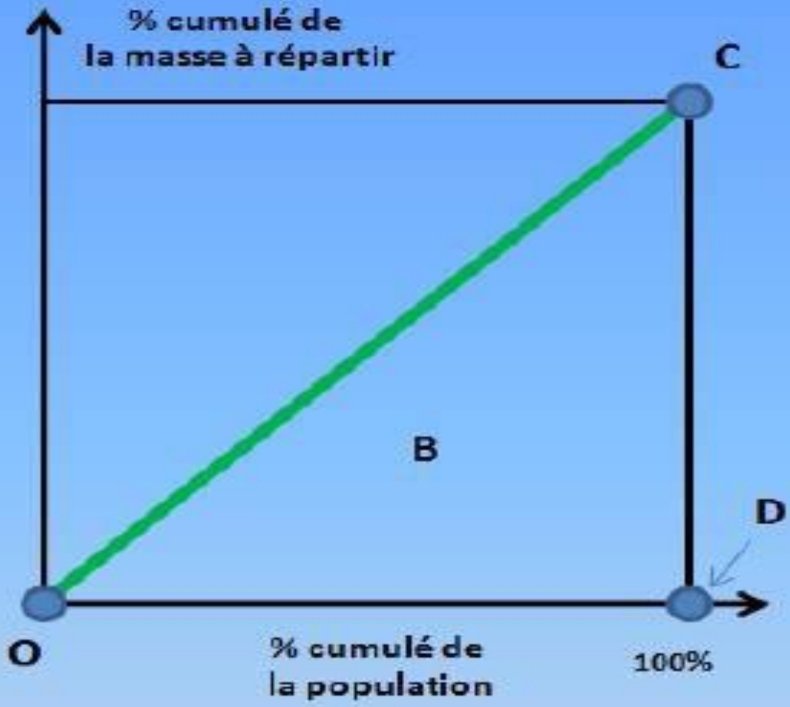
La courbe de Gini-Lorenz: L'indice de Gini

Exemple :calculer l'indice de Gini

L'indicede Gini est = **1** - 0.93166562 = **0.068334** → **La concentration est faible**

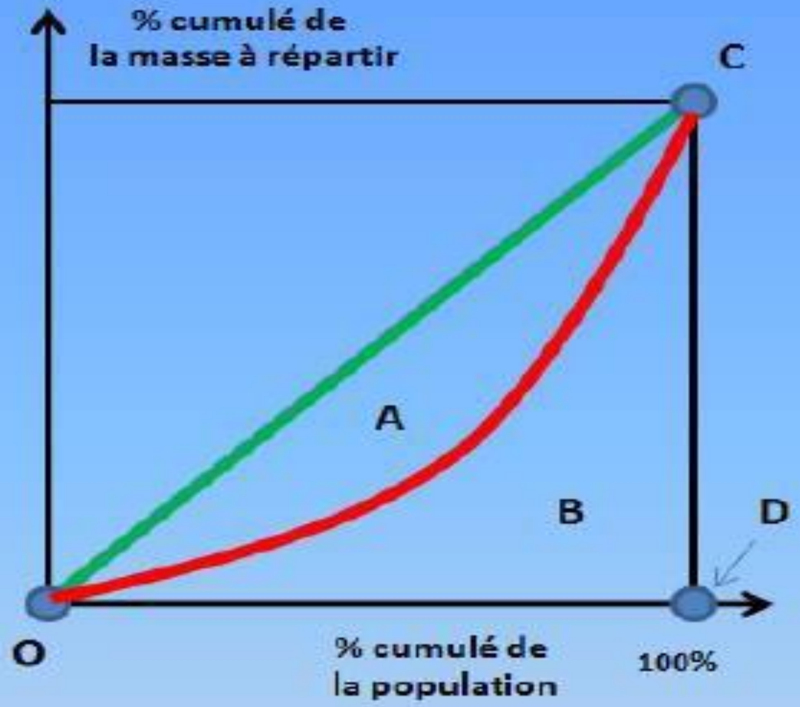


Cas numéro 1



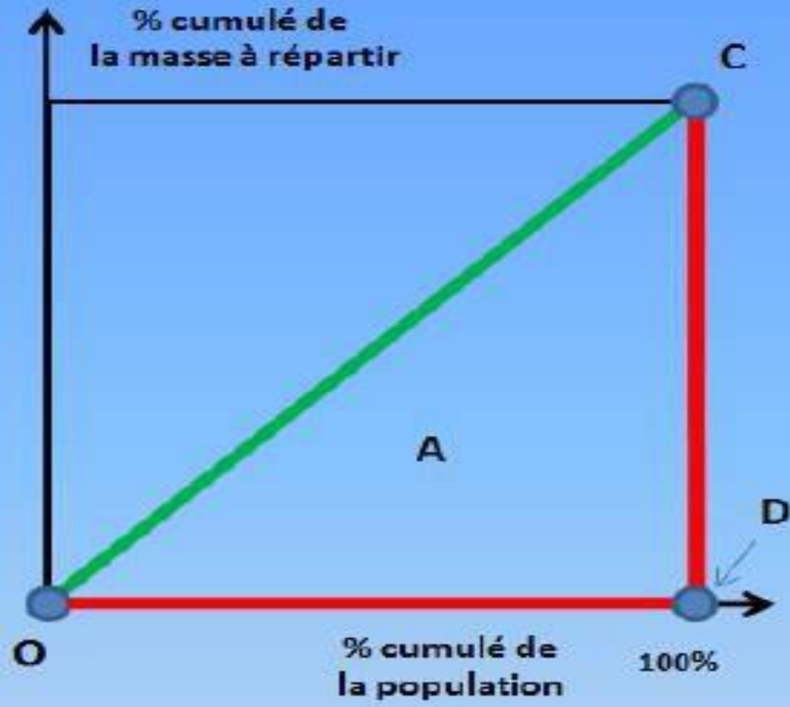
Egalité parfaite : la courbe de LORENZ se confond avec la droite OC d'égalité parfaite. Chaque individu de la population possède la même part de la masse totale

Cas numéro 2



Inégalité modérée: la courbe de LORENZ partage le triangle OCD est deux surfaces. Plus la surface A augmente aux dépends de la surface B et plus l'inégalité augmente

Cas numéro 3



Inégalité totale: la courbe de LORENZ est donné OCD. La surface A occupe tout le triangle OCD et la surface B a disparu. C'est le cas théorique où un seul individu possède 100% de la masse totale et les autres rien.

Fin du programme

Merci