Les caractéristiques de concentration

La mesure de la concentration revient à celle de la conséquence de la dispersion. Très importante en économique (concentration des salaires, des revenus, de la taille des entreprise…). Elle concerne des variables continues ne pouvant prendre que des valeurs positives.
C'est la conséquence de la dispersion ; On la détermine par 2 méthodes : le calcul et les graphes.

Paragraphe 1 : Détermination par le calcul

La démarche est la suivante :

1) On calcule la médiane (Mé) de la série
2) On calcule la médiale (Ml) de la série
3) On mesure l'écart (Δ M) entre la médiale et la médiane
4) On compare cet écart (Δ M) à l'intervalle de variation de la série.

La médiale est une médiane que l'on calcule non plus sur les effectifs (ni) de la série mais sur le produit ni xi (xi étant le centre de classe).

Paragraphe 2 : Détermination par le graphe

Cette analyse a été développée par l'italien Corrado Gini au cours de ses travaux sur les disparités de revenus et a abouti à la construction d'une courbe dite « de concentration » et à la détermination d'un ratio : l'indice de Gini

A) La courbe de concentration

Elle se construit sur un repère orthonormé à partir des fréquences cumulées relatives :

- Sur l'axe des abscisses : on reporte les valeurs de F(x) (fonction de répartition), c'est-à-dire la colonne où l'on a fait la somme des fi. Les valeurs varient de 0 à 1.
- Sur l'axe des ordonnées : on reporte les valeurs de F(nx), c'est-à-dire la colonne ou l'on a fait la somme des $\frac{ni\,xi}{\sum ni\,xi}$ pour calculer la médiale Ml. Les valeurs varient de 0 à 1.

$$F(x) = \sum \frac{ni}{n} = \sum fi \quad \text{et} \quad F(nx) = \sum \frac{ni\,xi}{\sum ni\,xi}$$

On obtient donc le carré ABCD qui porte le nom de **« carré de Gini ».**

Calcul de la concentration des salaires

On donne la série suivante d'un échantillon de 43 locataires selon leurs classes de salaire net après paiement du loyer et des impôts, exprimés en 10^2 €.

Classes 10^2	Effectif n_i	Centre de classe x_i	f_i	$F(x)$		Masse salariale $n_i\, x_i$	$\frac{n_i x_i}{\Sigma n_i x_i}$	$\sum_1^i \frac{n_i x_i}{\Sigma n_i x_i}$	
[4 ; 6 [	5	5	0,11	0,11		25	0,06	0,06	
[6 ; 8 [	8	7	0,19	0,30		56	0,14	0,20	
[8 ; 10 [	12	9	0,28	0,58	0,5	108	0,27	0,47	
[10 ; 12 [	10	11	0,23	0,81		110	0,27	0,74	0,5
[12 ; 14 [	8	13	0,19	1,00		104	0,26	1,00	
	43		1			403	1		

1. – La médiane est déterminée par :

$$\frac{Mé - 800}{0,5 - 0,30} = \frac{1\,000 - 800}{0,58 - 0,30} \Rightarrow \boxed{Mé = 942,86\ €}$$

2. – La médiale est déterminée par :

$$\frac{Ml - 1\,000}{0,5 - 0,47} = \frac{1200 - 1\,000}{0,74 - 0,47} \Rightarrow \boxed{Ml = 1\,022,22\ €}$$

3. – L'écart entre la médiale et la médiane est :

$$\Delta M = Ml - Mé = 1\,022,22 - 942,86 = \boxed{79,36\ €}$$

Cet écart ΔM traduit la concentration. On le compare **généralement à l'étendue** ou intervalle de variation de la série :

4. – Mesure de la concentration :

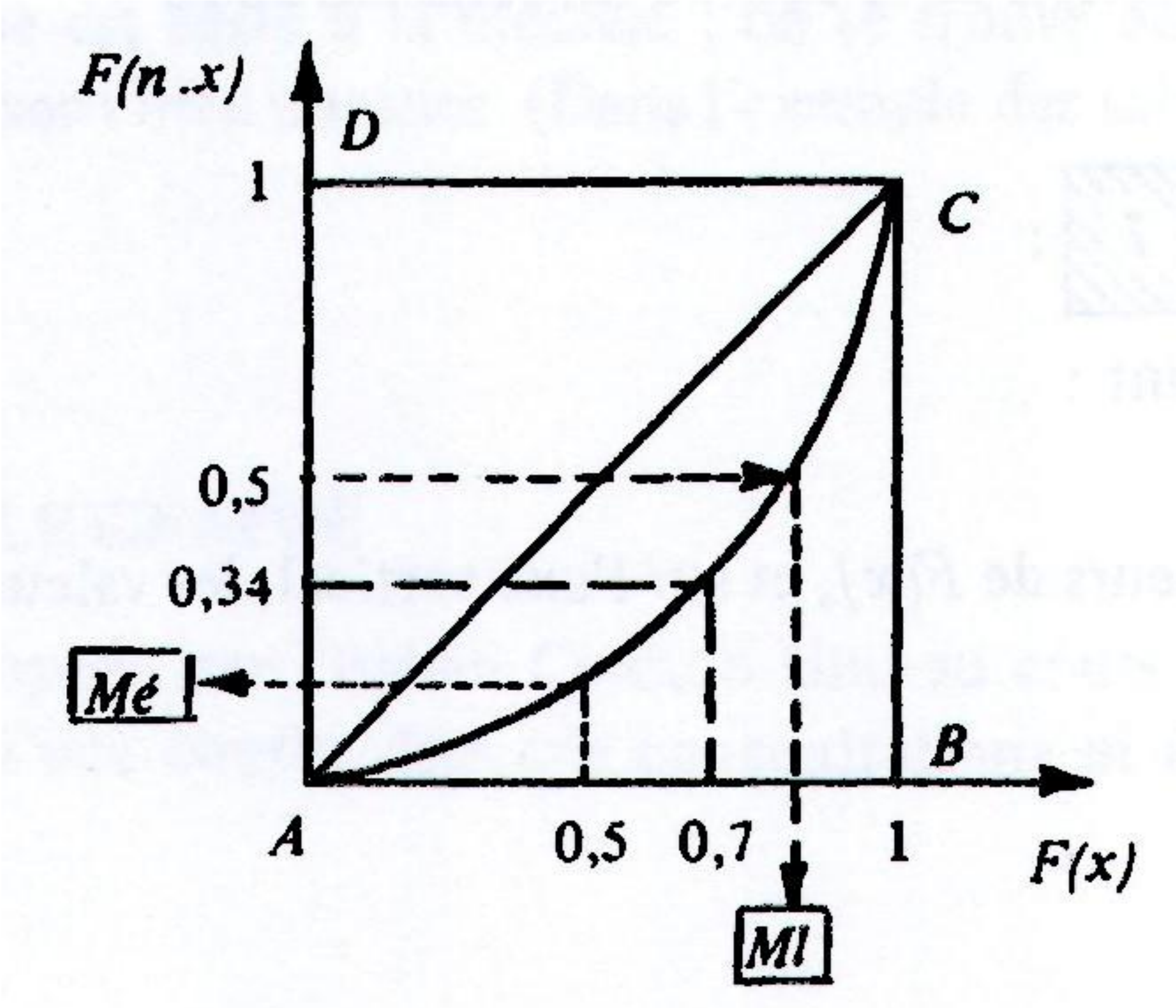
L'intervalle de variation est : 1 400 € – 400 € = 1 000 €.

$$\frac{\Delta M}{1\,000} = \frac{79,36}{1\,000} = \boxed{0,079}$$

La concentration des salaires est faible (de l'ordre de 8 %).

La concentration est forte si elle dépasse les 15%.

On construit la courbe de concentration (appelée aussi courbe de Lorenz, point par point : chaque point de la courbe a pour abscisse une valeur de F(x) et pour ordonnée la valeur de la fréquence cumulée relative de la totalité du phénomène (ni xi) correspondante.



Dans le schéma ci-dessus, F(x) = 0,7 pour F(nx)=0,34 ; si l'on reprenait l'exemple des salaires, on pourrait dire que 70% des salariés se partagent 34% de la masse salariale. La bissectrice AC correspond à la ligne ***d'équirépartition parfaite***, par construction. Toujours dans l'exemple des salaires, 10% des salariés recevraient 10% de la masse salariale ; 20%, etc…) C'est la ligne de concentration nulle.

Donc : plus *la courbe de concentration s'écarte de la bissectrice, plus la concentration est forte*.
Ou encore : *plus la concentration est forte, plus l'aire de concentration se rapproche de l'axe des abscisses.*
C'est le message visuel.

B) Indice de Gini

C’est un ratio qui permet des comparaisons.

Il est égal au rapport de deux surfaces :

- Au numérateur, on porte la surface comprise entre la bissectrice et la courbe de concentration. Cette surface prend le nom de ***surface de concentration***.
- Au dénominateur on porte la surface du triangle ABC.

$$I_G = \frac{Aire\ de\ concentration}{Aire\ du\ triangle\ ABC} = 2 \times \text{Aire de concentration}$$

I_G varie de 0 à 1 (d’une concentration nulle à maximale).

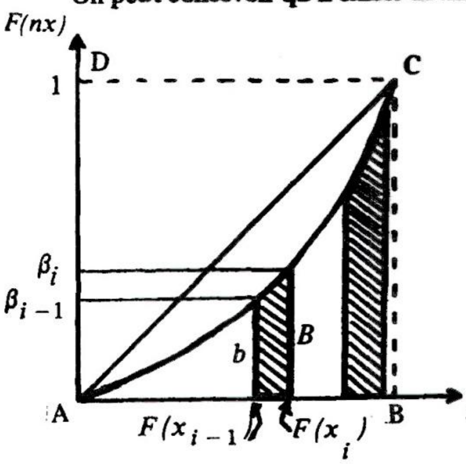
car Aire du Triangle ABC = $\frac{1}{2}$

Le problème est de mesurer les aires sans avoir recours au calcul intégral.

Si l’on tient à calculer une valeur numérique de I_G, on peut se servir de la méthode des trapèzes.

METHODE DES TRAPEZES

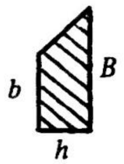
On peut concevoir qu’il existe autant de trapèzes que de classes, comme le montre la figure ci-dessous :



Donc β_i est la valeur de $F(n.x)$ de la ligne « i » du tableau β_{i-1} est la valeur précédente.($\beta_{i-1} = 0$ pour la valeur $i = 1$).

En règle générale : $\beta_i = \sum_{h=1}^{i} \frac{n_i x_i}{\sum_i n_i x_i}$

– Rappelons que la surface d’un trapèze est donnée par :



$$S = \frac{(b + B)h}{2}$$

Dès lors, l’aire de concentration est égale à l’aire du triangle ABC moins la somme des aires des trapèzes, soit :

$$\text{Aire de concentration} = \frac{1}{2} - \Sigma \frac{(b+B)h}{2} = \frac{1}{2} - \frac{1}{2} \Sigma [\beta_{i-1} + \beta_i] \cdot [F(x_i) - F(x_{i-1})]$$

$$= \frac{1}{2} - \frac{1}{2} \Sigma (\beta_{i-1} + \beta_i) f_i$$

Et , $I_G = 2 \times$ Aire de concentration $\Rightarrow$ $\boxed{I_G = 1 - \Sigma(\beta_{i-1} + \beta_i) f_i}$

Il suffit de disposer les calculs comme suit :

β_{i-1}	β_i (F(nx))	$\beta_{i-1}+\beta_i$	$f_i(\beta_{i-1}+\beta_i)$
0	–	–	–
–	–	–	–
–	–	–	–
			Σ

$I_G = 1 - Σ$

I_G est donné par : 1 moins la somme des termes de la dernière colonne.