Université A- mira de Bejaia
Faculté des sciences économiques, commerciales et des sciences de gestion
Département SEGC(LMD)
Module stat I
Enseignant Dr. Mousli

Chapitre 6 : Les paramètres *de concentration*

Introduction :

La mesure de la concentration revient à celle de la conséquence de la dispersion. Très importante en économie (concentration des salaires, des revenus, de la taille des entreprise…). Elle concerne des variables continues ne pouvant prendre que des valeurs positives.

Il existe deux méthodes de détermination de la concentration : par le calcul et par les graphes.

6.1. Détermination de la concentration par le calcul

La démarche est la suivante :

1) On calcule la médiane (Me) de la série
2) On calcule la médiale (Ml) que nous définissons plus bas.
3) On mesure l'écart (ΔM) entre la médiale et la médiane.
4) On compare cet écart (ΔM) à l'étendue.

6.1.1. Détermination de la médiane (Me)

Nous savons effectuer ce calcul (voir chapitre 3), qui passe par la résolution d'une interpolation linéaire.

6.1.2. Détermination de la médiale (Ml)

Définition pratique :

La médiale de la série (c_i, n_i) est la médiane de la série $(c_i\ ; n_i c_i)$. (c_i étant le centre de classe).

La médiale est la valeur du caractère étudié qui partage donc la série $(c_i\ ; n_i c_i)$ en deux sous-ensembles égaux. C'est une caractéristique de valeur centrale.

Le principe du calcul de la médiale est exactement le même que celui de la médiane, mis à part le fait qu'on l'applique, non plus sur la colonne des fréquences relatives cumulées $f_i \uparrow$ mais sur celle des fréquences relatives cumulées des $f_i' \uparrow$ avec : $\left(f_i' = \frac{n_i c_i}{\sum n_i c_i} \right)$

Exemple :

On donne le tableau suivant concernant un échantillon de 40 salariés d'une entreprise selon leurs classes de salaires horaires. D'abord, il faut former la colonne des $n_i c_i$ et celle des fréquences cumulées des $n_i c_i$ $\left(f_i' \uparrow\right)$. Le produit $n_i c_i$ (appelée valeur globale) représente la masse salariale.

Classes	n_i	Centre de Classe c_i	f_i $\left(\frac{n_i}{N}\right)$	$f_i \uparrow$	Masse salariale $n_i c_i$	f_i' $\left(\frac{n_i c_i}{\sum n_i c_i}\right)$	$f_i' \uparrow$
[10-20[	5	15	0,125	0,125	75	0,052	0,052
[20-30[	7	25	0,175	0,300	175	0,121	0,173
[30-40[	12	35	0,300	0,600	420	0,290	0,463
[40-50[	10	45	0,250	0,850	450	0,310	**0,773**
[50-60[	6	55	0,150	1,000	330	0,227	1,000
Total	40	-	1	-	1450	1	-

La classe médiale est la classe [40-50[

$$Ml = X_{\min} + \frac{0{,}5 - f'_{iMl-1} \uparrow}{f'_{iML}} \times ai$$

$$Ml = 40 + \frac{0{,}5 - 0{,}463}{0{,}31} \times 10 = 41{,}19$$

La médiale de la distribution des salaires est donc la valeur du salaire qui partage la masse salariale en deux sous-ensembles égaux : Dès lors, le salaire médial est tel que les salariés qui se situent en deçà, gagnent autant que les salariés qui se situent au-delà.

6.1.3. L'écart (médiale – médiane)

La médiale est supérieure à la médiane (puisqu'on raisonne « en masse » dans le premier cas).

$\Delta M = Ml - Me$. Le calcul est immédiat.

6.1.4. Comparaison de ΔM à l'Étendue.

- Si ΔM est grand par rapport à l'étendue, la concentration est forte (Dans l'exemple des salaires, cela signifierait que l'inégalité entre les salaires est forte).
- Si ΔM est petit par rapport à l'étendue, la concentration est faible (Dans l'exemple des salaires, cela signifierait qu'il n'y a pas de grandes disparités salariales entre les classes des salaires).
- Si ΔM est nul, la médiane est égale à la médiale ; on se trouve donc dans une situation d'égalité parfaite ou d'équi-répartition, si les classes sont bien choisies. (Dans l'exemple des salaires, tous les salariés toucheraient le même salaire).

Pour notre exemple, nous avons les calculs suivants :

$$Me = X_{\min} + \frac{0{,}5 - f_{iMe-1}\uparrow}{f_{iMe}} \times ai$$

$$\rightarrow Me = 30 + \frac{0{,}5 - 0{,}3}{0{,}3} \times 10 = 36{,}66$$

$$\Delta M = Ml - Me \rightarrow \Delta M = 41{,}19 - 36{,}66 = 4{,}53\,.$$

Cet écart ΔM traduit la concentration.

- **On calcule l'étendue** : $E = X_{\max} - X_{\min} = 60 - 10 = 50$
- **On compare** ΔM **à** E :

On remarque que ΔM est petit par rapport à l'étendue, donc la concentration des salaires est faible. Elle est de l'ordre de 9%.

$$\left(\frac{\Delta M}{E} = \frac{4{,}53}{50} = 0{,}0906 \approx 9\% \right)$$

6.2. Détermination de la concentration 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**.

Elle se construit sur un repère orthonormé à partir des fréquences.

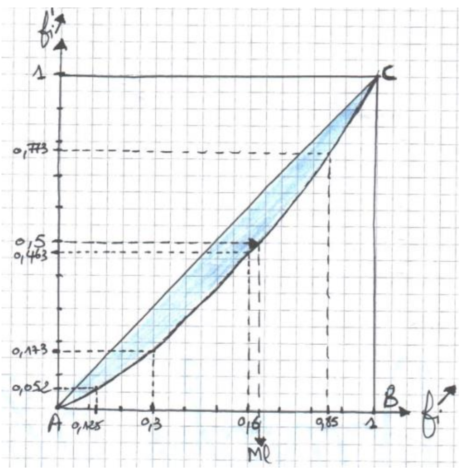
6.2.1. La courbe de concentration (ou de LORENZ)

C'est la courbe obtenue en représentant :

*En abscisse, on utilise les fréquences cumulées croissantes de la série (c_i ; n_i) $f_i\uparrow$

*En ordonnée, on utilise les fréquences cumulées croissantes de la série (c_i ; $n_i c_i$) $f_i'\uparrow$

Reprenons l'exemple précédent :



On voit donc que 30% des salariés se partagent seulement 17,3% de la masse salariale. La concentration reste néanmoins faible, si l'on en juge par la surface de l'aire grisée, de façon uniquement visuelle.

La diagonale (AC) correspond à la ligne d'**équirépartition parfaite** (la ligne de concentration nulle)**.** C'est-à-dire **:** 10% des salariés recevraient 10% de la masse salariale, 20%, etc.

6.2.2. L'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 diagonale et la courbe de concentration.

Au dénominateur, on porte la surface du triangle ABC.

$$I_G = \frac{\text{aire de concentration}}{\text{aire de triangle ABC}}$$

En effet : l'aire du triangle ABC est de $(1 \times 1)/2=0,5$; et diviser par 0,5, revient à multiplier par 2, donc **l'indice de Gini** ($I_G = 2 \times$ aire de concentration).

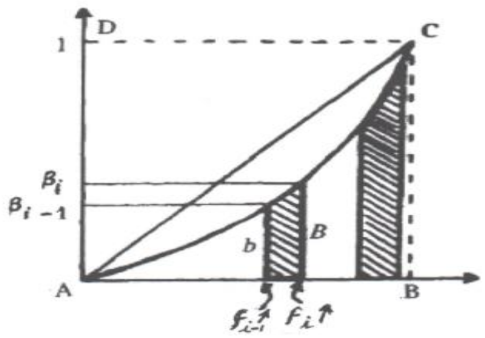
I_G Varie de 0 à 1(d'une concentration nulle à une concentration maximale).

Le problème est de mesurer les aires sans avoir recours au calcul intégral. Plusieurs méthodes graphiques sont possibles. La plus simple consiste à compter les carreaux sur le graphique que l'on aura soigneusement construit sur papier millimétré. Cependant la représentation graphique a essentiellement pour objectif de transmettre un message visuel. Elle n'est que la visualisation de la concentration mesurée par le calcul.

Néanmoins, si l'on tient absolument à calculer une valeur numérique de I_G , on peut se servir (entre autres méthodes d'approximation) de celle donnée par la méthode des « Trapèzes » :

Méthode des Trapèzes

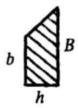
On peut concevoir qu'il existe autant de trapèzes que de classes, comme le montre la figure ci-dessus :



Donc β_i est la valeur de $f_i'\uparrow$ 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 : $\beta_i = \sum_{h=1}^{i} \frac{n_i c_i}{\sum_i n_i c_i}$

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 :

$$\text{Aire de concentration} = \frac{1}{2} - \sum \frac{(b+B)h}{2} = \frac{1}{2} - \frac{1}{2}\sum [\beta_{i-1} + \beta_i] \times [f_i\uparrow - f_{i-1}\uparrow]$$
$$= \frac{1}{2} - \frac{1}{2}\sum (\beta_{i-1} + \beta_i) \times f_i$$

Nous avons : $I_G = 2 \times \text{aire de concentration}$

Donc : $I_G = 2\left(\frac{1}{2} - \frac{1}{2}\sum (\beta_{i-1} + \beta_i) \times f_i\right) \Rightarrow I_G = 1 - \sum (\beta_{i-1} + \beta_i) \times f_i$

Il suffit de disposer les calculs comme suit :

β_{i-1}	β_i	$\beta_{i-1} + \beta_i$	f_i	$(\beta_{i-1} + \beta_i) \times f_i$
0	0,052	0,052	0,125	0,0065
0,052	0,173	0,225	0,175	0,039375
0,173	0,463	0,636	0,300	0,1908
0,463	0,773	1,236	0,250	0,309
0,773	1,000	1,773	0,150	0,26595
-	-	-	-	**0,811625**

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

Donc : $I_G = 1 - \sum(\beta_{i-1} + \beta_i) \times f_i \Rightarrow I_G = (1 - 0{,}811625) = 0{,}188375 \approx 18{,}84\%$

Exercice :

On donne le tableau suivant concernant un échantillon de 43 locataires selon leurs classes de salaires net en milliers de dinars (MDA) :

Salaires (MDA)	Nombre de locataires (n_i)
[40 - 60[	5
[60 - 80[	8
[80 - 100[	12
[100-120[	10
[120-140[	8
Total	43

1- Calculer la médiane
2- Calculer la médiale
3- Calculer l'écart ΔM (médiale – médiane)
4- Comment elle est la mesure de concentration des salaires ?
5- Calculer l'indice de Gini et interpréter le résultat.

Corrigé de l'exercice :

On forme d'abord la colonne de la masse salariale ($n_i c_i$) et celle des fréquences cumulées des $n_i c_i$ $(f_i' \uparrow)$

Salaires (MDA)	n_i	Centre de Classe c_i	f_i $\left(\frac{n_i}{N}\right)$	$f_i \uparrow$	Masse salariale $n_i c_i$	f_i' $\left(\frac{n_i c_i}{\sum n_i c_i}\right)$	$f_i' \uparrow$
[40 - 60[	5	50	0,11	0,11	250	0,06	0,06
[60 - 80[	8	70	0,19	0,30	560	0,14	0,20
[80 - 100[	12	90	0,28	0,58	1080	0,27	0,47
[100-120[	10	110	0,23	0,81	1100	0,27	**0,74**
[120-140[	8	130	0,19	1,00	1040	0,26	1,00
Total	43	-	1	-	4030	1	-

1) **Calcul de la médiane**

La classe médiane est la classe [80-100[

$$Me = X_{\min} + \frac{0{,}5 - f_{iMe-1}\uparrow}{f_{iMe}} \times ai \qquad Me = 80 + \frac{0{,}5 - 0{,}30}{0{,}28} \times 10 = 94286\text{DA}$$

2) **Calcul de la médiale**

La classe médiale est la classe [100-120[

$$Ml = X_{\min} + \frac{0{,}5 - f'_{iMl-1}\uparrow}{f'_{iML}} \times ai \qquad Ml = 40 + \frac{0{,}5 - 0.47}{0.27} \times 10 = 102222\text{ DA}$$

3) **Calcul de l'écart ΔM (médiale – médiane)**

$$\Delta M = Ml - Me = 102222 - 94286 = 7936\text{ DA}$$

Cet écart ΔM traduit la concentration. On le compare à l'étendue.

4) **Mesure de la concentration**

Etendue : $E = X_{\max} - X_{\min} = 140000 - 40000 = 100000\text{ DA}$

On remarque que : ΔM est petit par rapport à l'étendue, donc la concentration est faible. Dans notre exemple, nous avons $\frac{\Delta M}{E} = \frac{7936}{1000} = 0{,}079 \approx 8\%$

Comme la concentration des salaires est l'ordre de 8%, donc elle est faible.

5) **Calcul de l'indice de Gini**

Pour calculer l'indice de Gini, il suffit de disposer les calculs comme suit

β_{i-1}	β_i	$\beta_{i-1}+\beta_i$	f_i	$(\beta_{i-1}+\beta_i)\times f_i$
0	0,06	0,06	0,11	0,0066
0,06	0,20	0,26	0,19	0,0494
0,20	0,47	0,67	0,28	0,1876
0,47	0,74	1,21	0,23	0,2783
0,74	1,00	1,74	0,19	0,3306
-	-	-	-	0,8525

$I_G = 2\times$ **aire de concentration**

$$I_G = 1 - \sum(\beta_{i-1}+\beta_i)\times f_i$$

$$I_G = 1 - 0{,}8525 = 0{,}1475$$

$I_G = 14{,}75\%$, la concentration est néanmoins faible.