

CONCENTRATION

- Courbe de concentration (de Gini)
- Indice de concentration (de Gini)
- Médiale

REPÈRES

Il est des grandeurs pour lesquelles l'addition des observations n'a guère de sens : les taux d'alphabétisation, les températures, les notes en anglais, etc.

A l'inverse, les impôts sur le revenu payés par chacun s'ajoutent pour donner le montant total perçu à ce titre par l'État ; la somme des superficies des exploitations agricoles est la superficie agricole totale ; les productions nationales de voitures se cumulent dans la production mondiale, etc.

Dans le cas de grandeurs pour lesquelles le cumul a un sens, on définit la notion de *concentration*.

1. Courbe de concentration

a) Séries a_i et b_i

Deux cas se présentent, selon que les données initiales sont regroupées ou non.

● Série simple

Soit une série x_1, x_2, \dots, x_N , d'observations cumulables classées par ordre croissant.

Pour fixer les idées, on peut supposer qu'il s'agit de la production laitière des exploitations agricoles d'une région.

L'effectif total est N , tandis que la production totale, c'est-à-dire la masse globale du caractère, est :

$$P = x_1 + x_2 + \dots + x_N = \sum x_i.$$

On définit les séries de rapports : a_1, a_2, \dots, a_k et b_1, b_2, \dots, b_k .

Les a_i sont les parts cumulées successives de l'effectif total, c'est-à-dire les fréquences cumulées associées à la série simple des observations x_i :

$$a_1 = \frac{1}{N}, a_2 = \frac{2}{N}, \dots, a_i = \frac{i}{N}, \dots, a_N = \frac{N}{N} = 1.$$

Les b_i sont les parts cumulées successives de la masse globale du caractère, soit ici de la production totale :

$$b_1 = \frac{x_1}{P}$$

$$b_2 = \frac{x_1 + x_2}{P}$$

$$b_i = \frac{x_1 + x_2 + \dots + x_i}{P}$$

$$b_N = \frac{x_1 + x_2 + \dots + x_N}{P} = \frac{P}{P} = 1.$$

● Série regroupée

Dans le cas fréquent d'une série regroupée par valeurs croissantes : x_1, x_2, \dots, x_k , du caractère cumulable, avec les effectifs correspondants : n_1, n_2, \dots, n_k , on transpose aisément ce qui précède.

L'effectif total est :

$$N = n_1 + n_2 + \dots + n_k = \sum n_i$$

et la production totale :

$$P = n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_k \cdot x_k = \sum n_i x_i.$$

Les a_i sont les fréquences cumulées, soit ici :

$$a_1 = \frac{n_1}{N} = fc_1$$

$$a_2 = \frac{n_1 + n_2}{N} = fc_2$$

$$a_i = \frac{n_1 + n_2 + \dots + n_i}{N} = fc_i$$

$$a_k = \frac{n_1 + n_2 + \dots + n_k}{N} = \frac{N}{N} = 1 = f c_k.$$

Et les b_i sont les parts cumulées successives de la production totale, soit :

$$b_1 = \frac{n_1 \cdot x_1}{P}$$

$$b_2 = \frac{n_1 \cdot x_1 + n_2 \cdot x_2}{P}$$

$$b_i = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_i \cdot x_i}{P}$$

$$b_k = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_k \cdot x_k}{P} = \frac{P}{P} = 1.$$

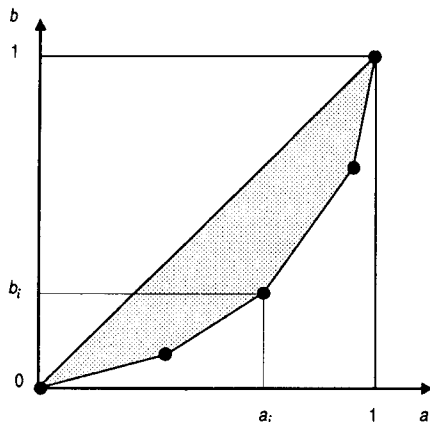
b) Construction de la courbe de concentration

Dans les deux cas, la courbe de concentration, ou encore courbe de Gini, est la courbe (ou plus exactement la suite de segments) qui, dans un repère $a \times b$, joint l'origine $(0,0)$ puis les points successifs (a_1, b_1) , (a_2, b_2) , ..., (a_k, b_k) , le dernier point étant donc le point $(1,1)$.

La courbe s'inscrit dans un carré de côté égal à l'unité, que l'on a coutume de représenter avec sa diagonale joignant les points $(0,0)$ et $(1,1)$.

On montre ainsi que les segments successifs sont de pente croissante et que la courbe est entièrement située sous la diagonale ou confondue avec elle (cas limite).

Figure 1



L'écart entre la courbe et la diagonale (surface grisée, fig. 1) traduit l'importance de la concentration. Plus la courbe est proche de la diagonale, plus la concentration est faible, plus elle s'en éloigne et plus la concentration est forte.

Il s'agit de la concentration au sens usuel du terme, c'est-à-dire, pour reprendre l'exemple de la production laitière, le fait qu'une petite partie des producteurs assure une grande part de la production.

2. Indice de concentration - Médiale

a) Indice de concentration

L'éloignement entre la courbe et la diagonale traduisant l'importance de la concentration, on définit, pour résumer cette information, l'indice de concentration ou *indice de Gini*, égal à deux fois la surface entre la courbe et la diagonale.

L'indice ainsi construit est toujours compris entre 0 et 1 :

- proche de 0, il indique une faible concentration ;
- proche de 1, il indique une forte concentration.

b) Médiale

La *médiale* est un concept analogue à celui de médiane, mais appliqué à un ensemble différent. On avait défini la médiane comme la valeur de l'observation qui voit les fréquences cumulées, a_i , dépasser 0,50 ; de même, la médiale est la valeur de l'observation qui voit les parts cumulées de la production, b_i , dépasser 0,50.

Les observations de plus fort poids étant cumulées en dernier, la médiale est toujours supérieure ou égale à la médiane.

L'écart entre la médiale et la médiane, rapporté à l'échelle des observations, donne une autre mesure de la concentration. Plus la médiale s'écarte de la médiane, plus la concentration est forte.

3. Utilisation de l'indice de concentration

L'étude de la concentration ne se limite pas aux données statistiques de production.

Toute autre série cumulative peut s'examiner sous cet angle : les revenus ou les impôts des particuliers, les stocks (en valeur par article) d'une entreprise, les capacités d'hébergement (en nombre de chambres d'hôtel) par régions, etc.

Outre l'examen d'une série isolée, la courbe et l'indice de concentration permettent la comparaison de plusieurs séries. Par exemple : les salaires en France et en Allemagne, ou encore la distribution alimentaire selon la taille des commerces en France en 1958 et en 1988, etc.

Appliqué par exemple à la distribution des revenus, l'indice de Gini revêt une signification particulière : il quantifie le degré d'inégalité de la répartition et autorise, à cet égard, des comparaisons dans le temps et dans l'espace.