CHAPITRE 2

INDICATEURS DE LA TENDANCE CENTRALE

→ *Mode et classe modale*
→ *Médiane et classe médiane*
→ *Moyennes*

REPÈRES

Si la répartition en fréquences recèle toute l'information statistique contenue dans la série des observations, il est souvent souhaitable de résumer celle-ci en quelques caractéristiques simples.

La première caractéristique à considérer, dans l'examen d'une série statistique, est sa *tendance centrale*. Les trois indicateurs présentés quantifient cette notion.

1. Mode - Classe modale

a) Mode

Soit une série statistique prenant des valeurs isolées (couleur des yeux, CSP, nombre d'enfants...), le *mode* est la valeur la plus fréquente. Il est noté : Mo.

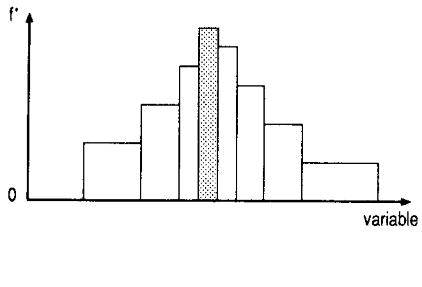
Exemple. Le mode de la série statistique relevant la couleur des yeux des Français est le marron.

b) Classe modale

Soit une série statistique numérique regroupée en classes, la *classe modale* est la classe de fréquence la plus élevée (si les classes sont d'amplitudes inégales, il s'agira de la classe de fréquence corrigée la plus élevée).

Une série statistique peut avoir plusieurs modes, ou classes modales. Le mode ou la classe modale désigne l'endroit où la répartition est la plus dense et correspond à la partie la plus haute du diagramme de fréquences (fig. 1).

Figure 1



2. Médiane - Classe médiane

a) Médiane

Soit une série statistique numérique ordonnée par valeurs croissantes, la *médiane* est la valeur de l'observation centrale, c'est-à-dire la valeur numérique telle qu'il y ait 50 % des observations qui lui soient inférieures et 50 % qui lui soient supérieures. Elle est souvent notée : Me.

Exemple. La médiane de la série : 1, 1, 1, 2, 2, 3, 4, 5, 8, 10, 13 est 3.

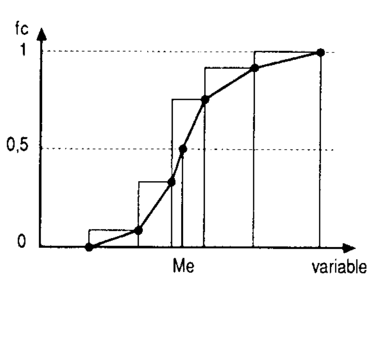
Pour des séries d'effectif pair, on convient de prendre la moyenne des deux observations centrales.

b) Classe médiane

Soit une série numérique regroupée en classes, la *classe médiane* est la classe qui contient l'observation centrale, ou médiane. En d'autres termes, c'est la première classe qui voit les fréquences cumulées atteindre 0,50 (ou 50 %).

Si, ne disposant pas des données initiales, on souhaite une valeur ponctuelle de la médiane, on fait une *interpolation* dans la classe médiane, c'est-à-dire qu'on calcule par proportionnalité le point de cette classe où le polygone de fréquences cumulées coupe la ligne horizontale : fc = 0,50 (fig. 2).

Figure 2



3. Moyenne arithmétique

La *moyenne arithmétique* est la plus usuelle. Elle est le plus souvent appelée simplement moyenne.

a) Moyenne d'une série simple

Soit une série de N observations numériques : x_1, x_2, ... , x_N. La moyenne arithmétique de ces N nombres, notée $\bar{x}$, est le nombre :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}.$$

Exemple. On achète trois litres de lait aux prix : 4,20 F, 4,30 F et 3,95 F. La moyenne est :

$$\frac{4{,}20 + 4{,}30 + 3{,}95}{3} = 4{,}15 \text{ (F)}.$$

Ce prix moyen est le prix unique qui aurait fait dépenser autant pour les trois litres.

b) Symbole sigma

En statistique, on est fréquemment amené à écrire des sommes telles que :

$$x_1 + x_2 + x_3 + \ldots + x_N$$
$$\text{ou bien : } x_1 \cdot y_1 + x_2 \cdot y_2 + \ldots + x_p \cdot y_p.$$

Il est alors commode d'utiliser une notation abrégée, le symbole *sigma* : Σ (de la lettre grecque sigma majuscule).

La somme $E_1 + E_2 + \ldots + E_p$ s'écrit ainsi :

$$\sum_{i=1}^{p} E_i$$

et, s'il n'y a pas d'ambiguïté sur l'ensemble des expressions $E_1, E_2, \ldots, E_p$ à additionner, on écrit simplement :

$$\sum_i E_i.$$

Le symbole Σ traduit donc une suite d'additions ; il conserve les propriétés de l'addition. En cas d'incertitude sur le sens d'une formule, il est facile de revenir à une écriture développée, avec des points de suspension.

La moyenne de la série simple : x_1, x_2, ... , x_N s'écrit donc plus simplement :

$$\bar{x} = \frac{\Sigma x_i}{N} = \frac{1}{N} \cdot \Sigma x_i .$$

c) Moyenne pondérée (ou avec coefficients)

Soit une suite de N nombres : x_1, x_2, ... , x_N auxquels sont associés les *poids*, ou *coefficients* : a_1, a_2, ... , a_N.

La moyenne des x_i *pondérée* par les a_i est le nombre :

$$\bar{x}_a = \frac{\Sigma a_i \cdot x_i}{\Sigma a_i}.$$

La moyenne pondérée n'est pas modifiée si tous les poids sont multipliés par un même nombre.

d) Moyenne d'une série regroupée

Soit une série statistique regroupée par valeurs : x_1, x_2, ... , x_k, observées avec les effectifs respectifs : n_1, n_2, ... , n_k. On voit que la moyenne des observations est :

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \ldots + n_k \cdot x_k}{n_1 + n_2 + \ldots + n_k}$$

$$\text{ou} \quad \bar{x} = \frac{\Sigma n_i \cdot x_i}{\Sigma n_i} = \frac{1}{N} \Sigma n_i \, x_i \ \text{(avec N} = \Sigma n_i\text{)},$$

$$\text{et puisque } f_i = \frac{n_i}{N}, \text{ on a : } \bar{x} = \sum_i f_i \, x_i .$$

En d'autres termes, la moyenne de la série est la moyenne des valeurs regroupées, pondérée par les effectifs correspondants.

Comme les fréquences sont proportionnelles aux effectifs, la moyenne de la série est aussi la moyenne des valeurs regroupées, pondérée par les fréquences, celles-ci étant de somme égale à 1.

4. Autres moyennes

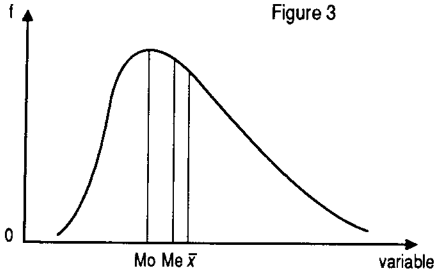
La moyenne arithmétique est la méthode la plus fréquemment utilisée pour définir une valeur « moyenne » d'observations d'une grandeur additive (par exemple, des poids, des sommes d'argent, des durées, etc.) ; il est cependant d'autres variables dont l'addition a moins de sens (la température, le coefficient intellectuel...).

Il est des grandeurs, enfin, qui se combinent autrement que par addition (par exemple, les taux d'inflation successifs, les vitesses sur différents parcours, etc.) Dans ces derniers cas, afin de déterminer des valeurs moyennes, il faut se reporter à la définition même des grandeurs. On est alors conduit à d'autres types de calculs.

5. Conclusions

On a défini trois indicateurs de la tendance centrale ; leurs valeurs sont en général différentes (exemple fig. 3 page suivante), sauf dans le cas de séries particulières (par exemple de distribution symétrique).

Figure 3



Leur interprétation et leur emploi sont également différents ; ainsi, pour une statistique de salaires :

– le salaire modal est le salaire le plus fréquent ;

– le salaire médian est celui du salarié qui voit autant de personnes gagner plus que de personnes gagner moins que lui ;

– le salaire moyen, enfin, est celui que chacun toucherait si tous les salaires étaient égaux.