Statistique : Caractéristiques d'une série statistique

Un caractéristique est une grandeur qu'on utilise pour résumer une série statistique.

On distingue deux sortes de caractéristique : caractéristiques de position et caractéristiques de dispersion.

1. Caractéristique de position

1.1 Le mode :

Le mode (ou dominante) est la valeur la plus fréquente de la variable.

C'est la variable qui a le plus grand effectif.

Le mode est défini même si la variable est qualitative.

Pour une série classée, dont les classes sont d'égal effectif, la classe modale est la classe qui correspond au plus grand effectif.

Si une série peut posséder un seul mode on dit qu'elle est unimodale. Si elle en possède plusieurs, on dit qu'elle est plurimodale.

1.2 La moyenne :

1.2.1 Définition :

La moyenne arithmétique d'une série statistique est égale à la somme des valeurs du caractère divisées par leur nombre.

i- Cas des données énumérées : $\bar{x} = \frac{x_1 + x_2 + \ldots + x_k}{N} = \frac{1}{N}\sum_{i=1}^{k} x_i$

ii- Si la série est donnée par sa distribution d'effectifs, les valeurs x_1, x_2, …, x_n ayant respectivement pour effectifs n_1, n_2, ..,n_k , alors :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{N} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$$

iii- Cas où les valeurs sont regroupées en classes : les ni valeurs de la i-ème classe sont supposées groupées au centre xi de la classe. On revient ainsi au cas précédent.

1.2.2 **Remarque** :

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

Donc si f_i est la fréquence de la variable x_i alors : $\bar{x} = \sum_{i=1}^{k} f_i x_i$

1.2.3 **Propriétés** (à établir en exercice) :

- Soit une série statistique sur une population et une partition de cette population en deux sous-populations d'effectifs respectifs n1 et n2. Si m1 et m2 sont les moyennes respectifs des deux sous-populations, alors la moyenne de la population est $\bar{x} = \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2}$

- Si $\bar{x}$ est la moyenne de la série (x_i, n_i) ,

alors la moyenne de la série $(x_i - a, n_i)$ est $\bar{x} - a$ et la moyenne de la série (hx_i, n_i) est $h.\bar{x}$

Donc , si on pose $y_i = ax_i + b$ alors la moyenne de la série (y_i, n_i) est $\bar{y} = a.\bar{x} + b$

1.3 La médiane :

1.3.1 Définition :

C'est la valeur de la variable qui partage la population en deux parties de même effectif : c'est donc la valeur M de xi telle que la moitié au plus des valeurs des xi soient inférieures à M et la moitié au plus des valeurs de xi supérieure à M.

1.3.2 Détermination de la médiane :

- Cas d'une série discrète

On range dans l'ordre croissant les valeurs de la variable, chaque valeur étant écrite autant de fois qu'elle est prise :

- si le nombre de valeurs est impair, la médiane est la valeur du milieu
- si le nombre de valeurs est pair, on peut prendre comme médiane tout valeur comprise entre les deux valeurs au centre. Par convention, on prend la demi somme de ces deux valeurs : $M = \frac{x_i + x_{i+1}}{2}$ si xi et xi+1 sont les valeurs au centre.

Exemple :

5 6 6 7 7 7 8 8 9 9 9 9 9 10 10 10 10 11 11 12 12 12 14 15 16

$\uparrow$
M

- Cas d'une série classée :

On trace le polygone des effectifs cumulés croissants et celui des effectifs cumulés décroissants. La médiane M est le l'abscisse du point d'intersection de ces deux courbes.

Un autre manière de la déterminer est de tracer le polygone des effectifs cumulés (ou fréquence cumulées) et la droite d'équation $y = \frac{N}{2}$ où N est l'effectif total de la population. La médiane est l'abscisse du point d'intersection de cette droite avec le polygone.

2. Caractéristiques de dispersion

Une caractéristique de dispersion est utilisée pour évaluer la dispersion d'une série. On utilise le plus souvent la variance et l'écart type .

2.1 Variance

2.1.1 Définition

La variance d'une série est la moyenne des carrés des écarts à la moyenne.

$$V=n_1(x-x_1)^2+n_2(x-x_2)^2+...+n_k(x-x_k)^2=\frac{\sum_{i=1}^{k} n_i\left(x-x_i\right)^2}{\sum_{i=1}^{k} n_i}$$

2.1.2 Remarques

- Plus la variance est grande, plus la série est dispersée.

Plus la variance est petite (voisin de 0), pus la série est resserrée autour de la moyenne.

- La variance est une quantité positive ou nulle.

2.1.3 Méthode de calcul

Même avec des valeurs observées x_i très simples, il arrive souvent que la moyenne $\bar{x}$ soit un nombre décimal. Dans ce cas, le calcul de la variance V nécessite des calculs fastidieux.

La formule de Koenig : $V=\frac{\sum_{i=1}^{k} n_i x_i^2}{N}-(\bar{x})^2$ (que vous allez démontrer en exercice) permet de simplifier les calculs.

2.2 Ecart type

L'écart type d'une série est la moyenne quadratique des écarts à la moyenne. C'est la racine carrée de la variance : $\sigma = \sqrt{V}$

Propriétés (à établir en exercice)

- La variance et l'écart-type de la série (x_i-a, n_i) sont indépendants de a :
  ce sont respectivement la variance et l'écart-type de (x_i, n_i)
- Si la variance et l'écart-type de la série (x_i, n_i) sont respectivement V et σ , alors la série $(h.x_{i,} n_i)$ a pour variance $V'=h^2V$ et pour écart-type $\sigma'=|h|\sigma$