

Statistiques

I) Présentation et Vocabulaire de base

Toute étude statistique s'appuie sur des données. Dans le cas où ces données sont numériques (95% des cas), on distingue les données discrètes (qui prennent un nombre fini de valeurs : par ex, le nombre de voitures par famille en France) des données continues (qui prennent des valeurs quelconques : par ex, la taille des animaux d'un zoo).

- Dans le cas d'une série discrète, le nombre de fois où l'on retrouve la même valeur s'appelle l'effectif de cette valeur. Si cet effectif est exprimé en pourcentage, on parle alors de fréquence de cette valeur.
- Dans le cas d'une série continue, on répartit souvent les données par classes.

Le but des statistiques est d'analyser les données dont on dispose. Pour cela, on peut par exemple chercher à déterminer la moyenne ou la médiane de la série. De tels nombres permettent notamment de comparer plusieurs séries entre elles. On les appelle indicateurs statistiques ou paramètres statistiques. On distingue les indicateurs de position (qui proposent une valeur "centrale" de la série) et les indicateurs de dispersion (qui indiquent si la série est très regroupée autour de son "centre" ou non).

Ainsi, le mode d'une série (valeur qui a le plus grand effectif de la série) est un indicateur de position.

L'étendue de cette série (différence entre la plus grande et la plus petite valeur) est un indicateur de dispersion.

La moyenne et la médiane sont des indicateurs de position.

De plus, lorsque la série est trop importante (population d'un pays...), on est obligé de faire un sondage, c'est à dire de restreindre l'étude à un échantillon de cette série. Tout le problème est alors de choisir un échantillon vraiment représentatif (de taille suffisante et non biaisé) et d'évaluer l'erreur commise par rapport à une étude qui porterait sur l'ensemble de la série. (exemple des sondages électoraux)

II) Médiane

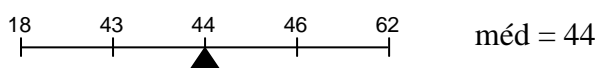
Définition

Soit une série statistique d'effectif total n , rangée par ordre croissant.

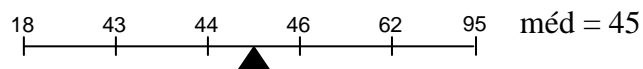
On appelle médiane la valeur "du milieu". On dit qu'elle partage la série en deux moitiés : il y a autant de valeurs en dessous qu'au dessus.

Pour déterminer son rang, il y a 2 cas :

- si n est impair : la médiane est la valeur de rang $\frac{n+1}{2}$



- si n est pair : nous prendrons la demi-somme des deux valeurs dont les rangs entourent le nombre $\frac{n+1}{2}$



Remarque :

Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. En revanche, on appellera classe médiane, la classe qui la contient (et permet donc d'en donner un encadrement).

Exemples

Données discrètes "en vrac"

10, 7, 12, 18, 16, 15, 5, 11, 11, 20, 15, 11, 18, 14

Ordonnons la série par ordre croissant : 5, 7, 10, 11, 11, 11, 12, 14, 15, 15, 16, 18, 18, 20

Il y a 14 termes or $\frac{14+1}{2} = 7,5$.

La médiane est donc la demi somme des 7^{ème} et 8^{ème} termes : méd = $\frac{12 + 14}{2} = 13$