

STATISTIQUES

www.mathsbook.fr

I - NOTIONS DE BASE

1 - VOCABULAIRE DE BASE

La **statistique** est tout simplement l'étude d'une **population** composée d'**individus**.

Le **caractère** : c'est l'aspect que l'on observe sur les individus. Il peut être qualitatif, quantitatif discret ou quantitatif continu.

Caractère qualitatif : Si l'on fait, par exemple, une étude statistique sur le mois de naissance d'une population, on parle de caractère qualitatif car on ne parle pas de valeurs numériques. En effet, les mois de l'année ne sont pas des valeurs numériques.

Caractère quantitatif : Si on fait au contraire une étude statistique sur l'âge d'une population, alors là (se sont des valeurs numériques) on parle de caractère quantitatif. On distingue deux caractères quantitatifs distincts :

- Discrète : 16 ans, 17 ans, 18 ans, etc.
- Continue : se sont tout simplement les intervalles : $[15; 20[$, $[20; 25[$, $[25; 30[$, etc.

2 - EFFECTIFS

Définition des effectifs : L'**effectif** de la valeur x_i est le nombre d'individus de la population ayant cette valeur ou appartenant à cette classe : on le note n_i .

L'**effectif total** N est la somme de tous les effectifs : $N = n_1 + n_2 + \dots + n_k$.

En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer l'**effectif cumulé croissant** en faisant la somme des effectifs de cette valeur et de tous ceux qui la précèdent.

3 - FRÉQUENCES

Définition des fréquences : la **fréquence** d'une valeur est le quotient de l'effectif de la valeur par l'effectif total.

En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer les **fréquences cumulées croissantes** en faisant la somme des fréquences de cette valeur et de tous ceux qui la précèdent.

4 - MÉDIANE

Définition de la médiane : C'est la valeur du caractère qui permet de partager la population N en deux groupes de même effectifs. On distingue deux cas : celui d'un caractère quantitatif discret et celui d'un caractère quantitatif continu.

Cas d'un caractère quantitatif discret :

- Si N est impair : la médiane est la valeur du caractère observé au rang $\frac{N+1}{2}$.
- Si N est pair : la médiane n'est pas définie, mais on convient de prendre pour médiane la moyenne des caractères observés au rang $\frac{N}{2}$ et $\frac{N}{2} + 1$.

Cas d'un caractère quantitatif continu : on construit la courbe des fréquences cumulées et la médiane est l'antécédent de 0,5.

5 - MOYENNE ARITHMÉTIQUE PONDÉRÉE

Définition de la moyenne arithmétique pondérée : La moyenne arithmétique pondérée, que l'on note \bar{x} , est donnée par la formule suivante :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n_1 + n_2 + \dots + n_k} = \frac{1}{N} \times \sum_{i=1}^{i=k} n_i x_i$$

Avec $N = n_1 + n_2 + \dots + n_k$ et n_i l'effectif de la valeur x_i .

II - PROPRIÉTÉS DE LA MOYENNE

Propriété de la moyenne : Soit une série statistique (x_1, x_2, \dots, x_n) de N éléments, réparties en p sous groupes ayant respectivement n_1, n_2, \dots, n_p éléments (tels que $n_1 + n_2 + \dots + n_p = N$) et de moyennes respectives $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p$.

Alors la série statistique a pour **moyenne** :

$$m = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + \dots + n_p\bar{y}_p}{N}$$

III - QUARTILE ET DÉCILE

Définitions : Soit une série statistique (x_1, x_2, \dots, x_n) de taille n .

- Les **quartiles** partagent la série statistique en quatre parties de même effectif.
Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des données soient inférieures ou égales à Q_1 .
Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des données soient inférieures ou égales à Q_3 .
- L'**intervalle interquartile** est l'intervalle $[Q_1; Q_3]$.
- L'**écart interquartile** est $Q_3 - Q_1$.
- Les **déciles** partagent la série en 10 parties de même effectif.