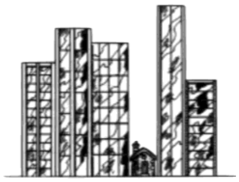


# STATISTIQUES



## I) Le vocabulaire utilisé en statistiques

### 1) Caractère d'une population

Les outils et les méthodes des études statistiques s'appliquent à des ensembles d'éléments nommés populations (exemple : l'ensemble des élèves d'un lycée, l'ensemble des pièces fabriquées en une heure par une machine, l'ensemble des trajets journaliers des élèves d'un LP).

Chaque élément de la population étudiée est : une unité statistique ou un individu (élève, pièce fabriquée, trajet journalier)

Le caractère ou variable statistique d'une population est la propriété sur laquelle porte l'étude statistique.

Le caractère statistique peut être :

- qualitatif (couleur d'une voiture, marque d'un appareil électroménager)

- quantitatif (Dans ce cas le caractère statistique est mesurable, il peut être noté par une variable statistique.)

Une variable statistique peut être :

- discrète (Elle prend un nombre fini de valeurs comme par exemple un nombre de voitures par puissance fiscale ou un nombre de personnes par foyer.)

- continue (Elle prend toutes les valeurs à l'intérieur d'un intervalle donné.)

L'étude statistique d'une population par rapport à une variable continue impose de regrouper le grand nombre de valeurs en tranches ou classes. (classes d'âge pour une population de personnes ; classes du montant des achats pour une population de clients d'une grande surface.)

### 2) Classes et effectifs

Une classe, c'est la portion de l'intervalle auquel appartiennent les valeurs de caractère.

Une série statistique associe à chaque valeur $x_i$ du caractère le nombre d'individus correspondant, appelé effectif partiel et noté $n_i$.

L'effectif de la population est noté $N$ et on a : $N = \sum_{i=1}^{i=p} n_i$

On appelle effectif cumulé croissant de la valeur $x_i$ (noté $n_i \nearrow$) la somme des effectifs de toutes les valeurs du caractère inférieures ou égales à $x_i$.

On appelle effectif cumulé décroissant de la valeur $x_i$ (noté $n_i \searrow$) la somme des effectifs de toutes les valeurs du caractère supérieures ou égales à $x_i$.

La fréquence d'une valeur $x_i$ du caractère est le quotient de l'effectif $n_i$ de ce caractère par l'effectif total $N$ : $f_i = \frac{n_i}{N}$

**Remarques** :
- La somme des fréquences est égale à 1. $\sum_{i=1}^{i=p} f_i = 1$
- Les fréquences sont souvent exprimées en pourcentage après multiplication par 100 du rapport $\frac{n_i}{N}$.

La fréquence cumulée croissante de la valeur $x_i$ (notée $f_i \nearrow$) est le rapport $\frac{n_i \nearrow}{N}$

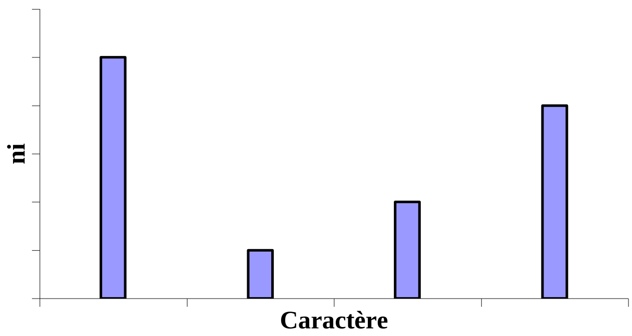
La fréquence cumulée décroissante de la valeur $x_i$ (notée $f_i \searrow$) est le rapport $\frac{n_i \searrow}{N}$

## II) Différentes représentations graphiques

### 1) Diagramme en bâtons

On l'utilise pour les séries à caractère discret. Pour celles qui utilisent un repère cartésien :
- sur l'axe des abscisses : valeur du caractère ;
- sur l'axe des ordonnées : valeurs des effectifs ou des fréquences.

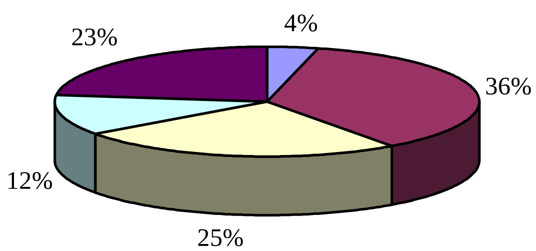
Principe : les hauteurs des différents bâtons sont proportionnelles aux effectifs correspondants.




### 2) Diagramme circulaire ou à secteurs

On l'utilise dans le cas d'une variable discrète.

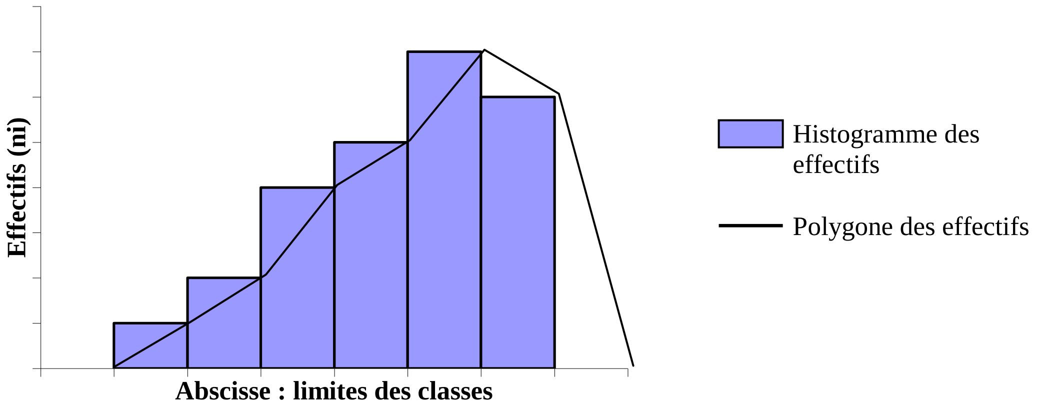
Principe : chaque secteur a un angle au centre de mesure proportionnelle à la fréquence de la classe correspondante exprimée en pourcentage.



## 3) Histogramme et polygone des effectifs

On l'utilise pour les séries à caractère continu, lorsque les valeurs de la variable sont réparties en classes.

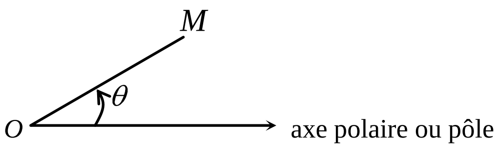
Principe : les aires des différents rectangles sont proportionnelles aux effectifs (aux fréquences) correspondantes.

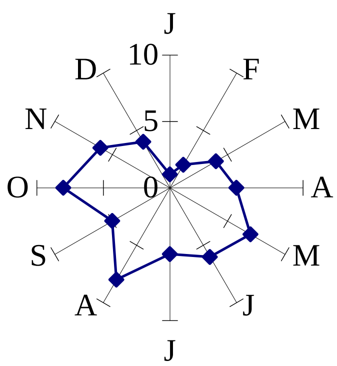



Remarque : lorsque les classes n'ont pas la même amplitude on prend un intervalle unitaire.

## 4) Diagramme polaire

On l'utilise pour les séries chronologiques.
Le repérage se fait :
- avec l'angle $\theta$
- avec la longueur $OM$.



# III) Caractères de position

## 1) Mode d'une série statistique

On appelle mode d'une série statistique à caractère discret la valeur du caractère statistique (notée $M_0$) qui correspond au plus grand effectif. (mode = dominante).

On appelle classe modale d'une série statistique à caractère continu la classe qui correspond au plus grand effectif.

Le mode est le centre de la classe modale.

## 2) Médiane d'une série statistique

C'est la valeur (notée $M_e$) de la variable pour laquelle il existe dans cette série autant de valeurs plus grandes que de valeurs plus petites.

La médiane se détermine graphiquement à l'aide du point d'intersection du polygone statistique des effectifs cumulés croissants et décroissants.

La médiane peut aussi être calculée dans le cas d'une série à caractère continu en utilisant la méthode de l'interpolation linéaire.

## 3) Quartiles

Les trois quartiles sont les trois valeurs du caractère qui partagent la population totale en quatre parties d'effectifs égaux.

Le premier quartile $Q_1$ correspond à 25 % de l'effectif total.

Le deuxième quartile $Q_2$ correspond à la médiane (50 % de l'effectif total).

Le troisième quartile $Q_3$ correspond à 75 % de l'effectif total.

L'intervalle interquartile est la différence entre les quartiles extrêmes ; il a pour valeur $Q_3 - Q_1$.

L'écart interquartile relatif : $\dfrac{Q_3 - Q_1}{Q_2}$

## 4) Déciles

$D_1, D_2, \ldots, D_9$ ; chaque décile partage en dix parties égales l'effectif total.

L'intervalle interdécile est la différence entre les déciles extrêmes : il a pour valeur $D_9 - D_1$

# IV) Caractères de dispersion

## 1) Calcul d'une moyenne d'une série distribuée en classes

On appelle moyenne d'une série statistique et on note $\overline{x}$ le nombre :

$$\overline{x} = \frac{\sum_{i=1}^{i=p} x_i n_i}{N} = \sum_{i=1}^{i=p} f_i x_i$$

$x_i$ désigne le centre de classe.

## 2) Étendue

L'étendue est la différence entre la plus grande valeur et la plus petite valeur de la série.

## 3) Écart moyen d'une série statistique

L'écart moyen est une caractéristique qui définit la dispersion des valeurs d'une série statistique. L'écart moyen est égal à la moyenne des écarts à la moyenne.

$$E_m = \frac{\sum_{i=1}^{i=p} n_i \left| x_i - \overline{x} \right|}{N}$$

$x_i$ : valeur ou centre de classe
$n_i$ : effectif correspondant
$N$ : effectif total
$\overline{x}$ : moyenne

## 4) Variance

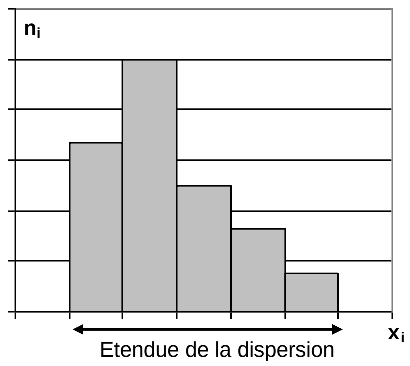
La variance $V$ est la moyenne des carrés des écarts à la moyenne.

$$V = \frac{1}{N} \sum_{i=1}^{i=p} n_i \left( x_i - \overline{x} \right)^2 = \left( \frac{1}{N} \sum_{i=1}^{i=p} n_i x_i^2 \right) - \overline{x}^2$$

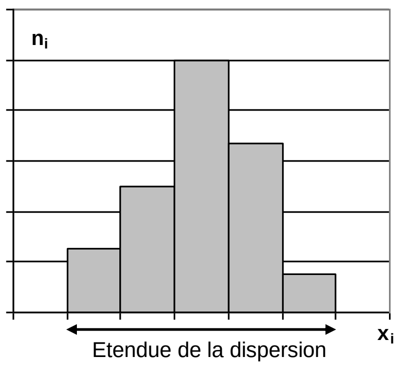
Avec $N$ : effectif total
$x_i$ : valeur de la variable
$n_i$ : effectif de la variable $x_i$
$\overline{x}$ : moyenne de la série

## 5) Écart-type

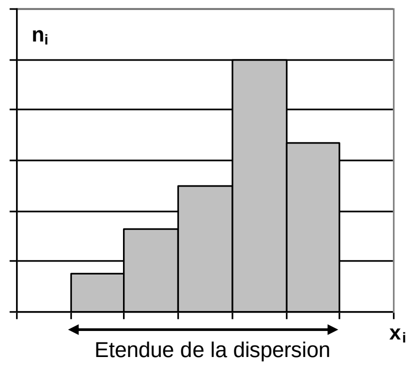
L'écart type définit la dispersion des valeurs d'une série statistique.



Les valeurs sont plus nombreuses vers la limite inférieure



Les valeurs sont réparties de part et d'autre de la classe au centre



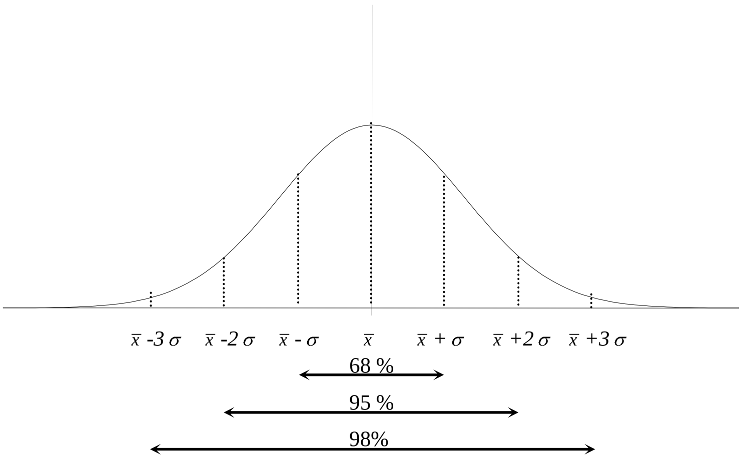
Les valeurs sont plus nombreuses vers la limite supérieure

L'écart-type $\sigma$ (lire : sigma) est la racine carrée de la variance : $\sigma = \sqrt{V}$

De nombreuses séries statistiques dont l'effectif est important ont une population distribuée suivant une loi dite normale avec une courbe des effectifs appelée courbe de Gauss.
Dans une loi normale, valeur moyenne, valeur médiane, valeur modale, sont égales.

Pour une série statistique « normalement » distribuée, il y a environ :

- 68 % de la population dans l'intervalle [ $\bar{x} - \sigma$ ; $\bar{x} + \sigma$ ]
- 95 % de la population dans l'intervalle [ $\bar{x} - 2\sigma$ ; $\bar{x} + 2\sigma$ ]
- 99% de la population dans l'intervalle [ $\bar{x} - 3\sigma$ ; $\bar{x} + 3\sigma$ ]




## 6) Coefficient de dispersion

On appelle coefficient de dispersion (exprimé en %) d'une série statistique de moyenne $\bar{x}$ et d'écart-type $\sigma$, le rapport $\frac{\sigma}{\bar{x}}$ (exprimé en % ; nombre abstrait)