

1.1 Statistique à une variable

Vocabulaire

Les études statistiques portent sur des ensembles appelés **populations** (ensemble des élèves d'un lycée, ...)

Chaque élément de la population est une **unité statistique** ou un **individu**.

Le **caractère** ou **variable statistique** est la propriété sur laquelle porte l'étude statistique.

Ce caractère peut être :

- Qualitatif (couleur ou marque d'une voiture, ...)
- Quantitatif (dans ce cas le caractère est mesurable)

Un caractère quantitatif peut être :

- Discret (il prend un nombre fini de valeurs)
- Continu (il prend toutes les valeurs dans un intervalle donné)

L'étude par rapport à une variable continue impose de regrouper le grand nombre de valeurs en intervalles ou classes.

La **variable statistique** ou le **centre de classe** est noté x_i .

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 $\frac{n_i}{N}$.

On appelle **mode** d'une série statistique à caractère discret la valeur du caractère statistique qui correspond au plus grand effectif.

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

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

Représentations graphiques

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.

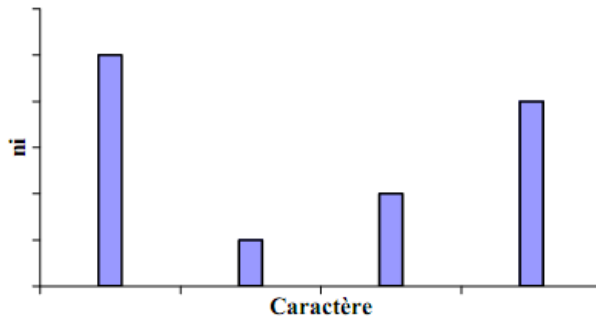
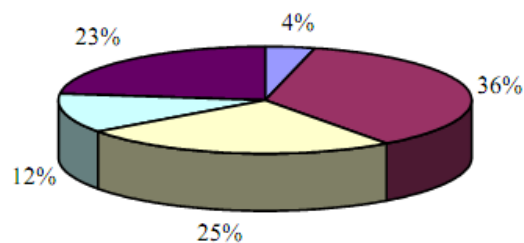


Diagramme circulaire ou à secteurs

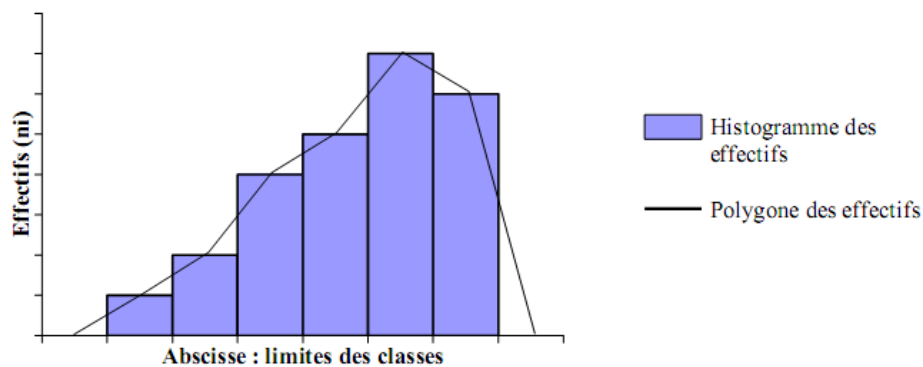
Principe : chaque secteur a un angle au centre de mesure proportionnelle à la fréquence de la classe correspondante exprimée en pourcentage.



Histogramme

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

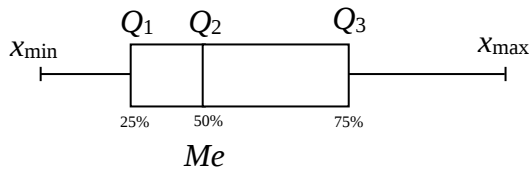
Principe : les aires des différents rectangles sont proportionnelles aux effectifs (aux fréquences) correspondantes.



Remarque : on peut aussi réaliser l'histogramme et le polygone des fréquences.

Diagrammes en boîte :

Il existe plusieurs types de diagrammes en boîte. La « boîte » est un rectangle limité par le premier et le troisième quartile et où figure la médiane. Les « moustaches » en revanche peuvent s'achever aux valeurs extrêmes, le minimum et le maximum de la série, ou aux premier et neuvième déciles, 10 % et 90 %, mais la notion de décile est hors programme.



Indicateur de tendance centrale

Moyenne

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

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

x_i désigne le centre de classe.

Médiane d'une série statistique

C'est la valeur (notée Me) 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.

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 $Q_3 - Q_1$.

Indicateurs de dispersion

Variance

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

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

Avec N : effectif total

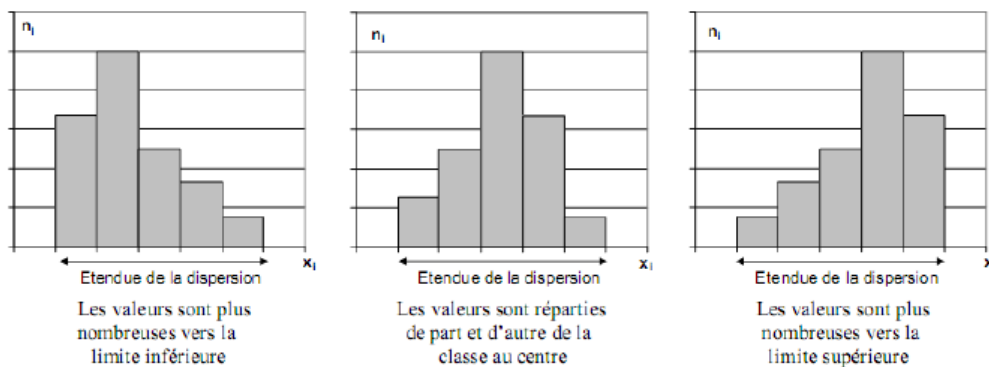
x_i : valeur de la variable

n_i : effectif de la variable x_i

$\bar{x}$: moyenne de la série

Ecart-type

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



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

On calcule d'abord la variance V

Puis on calcule l'écart-type σ en faisant la racine carrée de la variance V

Plus l'écart – type σ est grand, plus les valeurs du caractère sont dispersées autour de la moyenne

Plus il est petit, plus les valeurs du caractère sont groupées autour de la moyenne

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 $[\bar{x} - 3\sigma ; \bar{x} + 3\sigma]$

