Statistiques

I - Définitions

1 - La statistique :

La statistique est l'ensemble des méthodes qui ont pour objectif de collecter, de traiter et d'interpréter les données obtenues à partir d'une population constituée d'individus.

2 - Les différents type de variables :

Une variable peut être :

1. **Quantitative :** ce sont des variables à valeurs numériques (âges, poids, tailles, notes, etc...).
   On peut les classer en deux catégories : discrètes et continues.
   (a) *Discrète :* elles ne prennent qu'un nombre fini de valeurs (résultat d'un jet de dé, nombre d'enfant dans une famille).
   (b) *Continue :*elles prennent leurs valeurs dans un intervalle (par exemple entre 10 et 40 euros).
2. **Qualitative :** ce sont des variables représentées par des qualités. On peut citer par exemple la couleur des yeux, l'appréciation des clients pour un produit, le programme d ?études, etc.

II - Série statistique à une variable

Dans tout ce qui suit, on considère une série statistique de taille n, dont les termes (valeurs de la série) $x_1, x_2, ..., x_p$ sont rangés dans l'ordre croissant : $x_1 \leq x_2 \leq ... \leq x_p$.

Les valeurs $n_1, n_2, ..., n_p$ représentent respectivement les effectifs des termes $x_1, x_2, ..., x_p$.

On peut représenter toutes ces valeurs dans le tableau **(T)** suivant :

Valeurs de X	x_1	x_2	x_3	...	x_p
Effectifs	n_1	n_2	n_3	...	n_p

Exemple :

On donne dans les tableau ci-dessous le notes obtenues par 25 élèves d'une classe de Seconde en physique.

Notes	4	5	7	8	9	10	12	13	15	17	18
Effectifs	1	3	2	4	3	4	3	1	2	1	1

1 - Les paramètres statistiques :

a - La fréquence

La fréquence d'une valeur est égale à l'effectif de cette valeur divisé par l'effectif total.
Dans le cas du tableau **(T)**, la fréquence de la valeur x_i est égale à $f_i = \dfrac{n_i}{n}$.

La fréquence est donnée sur la troisième ligne du tableau ci-dessous.

Valeurs de X	x_1	x_2	x_3	...	x_p
Effectifs n_i	n_1	n_2	n_3	...	n_p
Fréquence f_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	...	$\frac{n_p}{n}$

b - La moyenne

La moyenne de la série, noté $\bar{x}$, est définie par : $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + ... + n_px_p}{n}$ ou $\bar{x} = f_1x_1 + f_2x_2 + ... + f_px_p$.

c - La variance

La variance de la série, notée V, peut être calculer à partirdes effectifs ou des fréquences :

- *Avec les effectifs :* $V = \dfrac{n_1(x_1-\bar{x})^2 + n_2(x_2-\bar{x})^2 + ... + n_p(x_p-\bar{x})^2}{n}$.

- *Avec les fréquences :* $V = f_1(x_1-\bar{x})^2 + f_2(x_2-\bar{x})^2 + ... + f_p(x_p-\bar{x})^2$.

d - L'écart-type

L'écart-type d'une série statistique, noté σ, est définie par : $\sigma = \sqrt{V}$. C'est la racine carrée de la variance.

Remarque : L'écart-type représente l'écart moyen des valeurs de la série par rapport à la moyenne. Une série statistique dont les valeurs sont très éloignées de la moyenne a un écart-type élevé.

e - La médiane

La médiane d'une série est la valeur séparant les valeurs d'une série en deux groupes de même effectif.

Au moins 50% des valeurs sont inférieures ou égales à la médiane.

Calcul de la médiane :

Premier cas :

Si l'effectif total n est pair, la médiane est la moyenne des deux termes centraux de la série : $M_e = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$.

Deuxième cas :

Si l'effectif n est impair, la médiane est le terme central de la série : $M_e = x_{\frac{n+1}{2}}$.

f - Les quartiles

- Premier quartile : Q_1

Le premier quartile, noté Q_1, est la plus petite valeur de la série pour laquelle au moins 25 % ($\frac{1}{4}$) des données lui sont inférieures ou égales.

- Troisième quartile : Q_3

Le troisième quartile, noté Q_3, est la plus petite valeur de la série pour laquelle au moins 75 % ($\frac{3}{4}$) des données lui sont inférieures ou égales.

g - Écart inter-quartile

On définie l'intervalle inter-quartile par : $[Q_1; Q_3]$.
On en déduit alors que l'écart inter-quartile est égal à : $Ec = Q_3 - Q_1$.

h - Étendue de la série

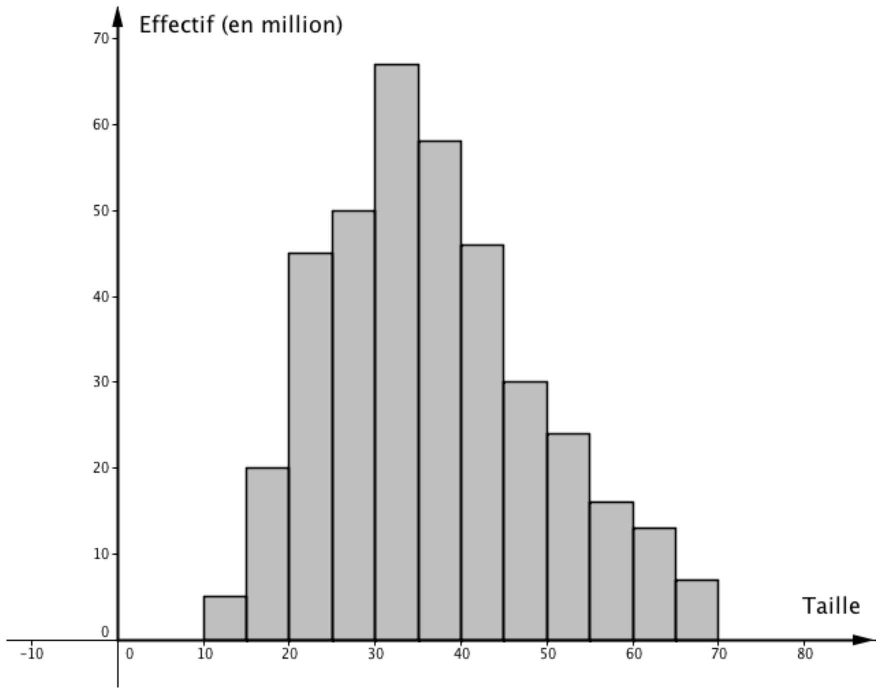
L'étendue de la série statistique est la différence entre la valeur maximale et la valeur minimale de la série.
On a : Étendue $= x_{max} - x_{min}$, avec x_{max} et x_{min} respectivement la valeur maximale et minimale.

II - Histogramme

L'histogramme est un graphique qui permet de représenter la répartition d"une variable continue.
Voici un exemple de la représentation d'un histogramme :

Taille	[40 ;45]	[45 ; 50]	[50 ;55]	[55 ; 60]	[60 ;65]	[65 ; 70]
Effectif (en million)	46	30	24	16	13	7

Taille	[10 ; 15]	[15 ; 20]	[20 ; 25]	[25 ;30]	[30 ; 35]	[35 ;40]
Effectif (en million)	5	20	45	50	67	58

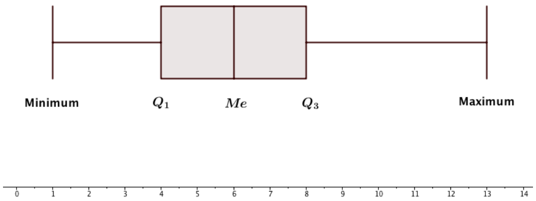


III - Diagramme en boîte

Le diagramme en boîte permet d'avoir une vue d'ensemble de la série statistique étudiée. Il est également appelé *boîte à moustache* ou encore le *diagramme de Tuckey* (nom de l'inventeur, 1977).

Méthode de construction :

Sur l'axe horizontal (ou vertical), on place (dans le même ordre) les valeurs suivantes : le minimum (x_{min}), le premier quartile (Q_1), la médiane (Me), le troisième quartile (Q_3) et le maximum (x_{max}).



Interprétation du diagramme en boîte :

— Si l'intervalle inter-quartiles est petit, les valeurs du "milieu" sont homogènes (proches). Dans le cas contraire les valeurs sont hétérogènes (dispersées).

— Plus les moustaches (l'étendue) sont grandes, plus les valeurs sont dispersées ; plus c'est petit, moins les valeurs sont dispersées.

— L'axe est orienté de gauche à droite. Plus la boîte est à gauche, plus les valeurs sont petites ; plus la boîte est à droite, plus les valeurs sont grandes.

— La symétrie du diagramme en boîte illustre une bonne répartition des valeurs de la série statistique.