

## Leçon 1:

### - La statistique descriptive

A. **Définition, nature et objet de la statistique**: Le mot "statistiques" au pluriel, désigne des collections de chiffres, présentés sous forme de tableaux parfois sous forme de graphiques, qui regroupent toutes les observations effectuées sur les faits nombreux, relatifs à un même phénomène.

Il est bien que qu'une pareille définition ne saurait suffire. Le statisticien, placé en face ces relevés chiffrés; devra les examiner de plus près, " de manière à obtenir des supports numériques sensiblement indépendants des anomalies du hasard et qui dénotent l'existence de causes régulières dont l'action est combinée avec celle des causes fortuites".

La statistique au singulier, c'est l'ensemble des procédés ou des méthodes qui auraient pour but l'étude mathématique des statistiques.

Si nous retenons ces deux définitions, du terme "statistique", nous dirons que la statistique est l'outil avec lequel on travaille une matière première constituée par les statistiques.

B. **Portée et limitation de l'emploi de la statistique**: La statistique aura pour but d'étudier des faits pour permettre ensuite de prendre des décisions. On conçoit alors qu'elle puisse être dangereuse si l'étude a été viciée par des erreurs d'observation; ou des erreurs d'interprétation.

Des critiques ont d'ailleurs été faites à la statistique:

- On lui a reproché de ne porter que sur des faits passés, et en conséquence, d'apporter trop tard des enseignements, mais cela n'est pas vrai pour les domaines où justement apparaissent des permanences statistiques.
- On a reproché aux statistiques d'être fausses; c'est vrai certaine mesure, mais il en est de toutes les bases numériques sur lesquelles ont été fondées les sciences expérimentales, la physique par exemple. Par ailleurs il est préférable d'avoir une connaissance même imparfaite des événements que pas des connaissances du tout.
- On a dit de la statistique qu'elle aboutisse à des conclusions relatives au comportement des ensembles et non celui des individus, mais précisément le statisticien ne s'intéresse qu'à des ensembles et non à des unités.
- On a accusé la statistique d'être une des formes de plus raffinées du mensonge. Un jugement sûr, et de la prudence, sont donc nécessaires en la matière avant de se livrer à toute conclusion. Il sera donc toujours indiqué d'établir soigneusement les statistiques qui seront utilisées, et de bien connaître les techniques qui serviront à leur étude.

C. **Concepts de base de la statistique**: D'une manière générale, la méthode statistique est basée sur quatre concepts: la population; les variables, les observations et données.

1. **Population et individu**: La notion fondamentale en statistique est celle d'ensemble d'objets équivalents que l'on appelle **population**, ses objets sont appelés **individus** ou **unités statistiques**.

Lorsque l'on observe qu'une partie de la population on parle de sondage.

La partie de la population étudiée est appelée **échantillon** et on cherche toujours à généraliser les résultats obtenus sur l'échantillon à toute la population.

- **Exemple1:**

- Etudier les notes de 1000 étudiants est une expérience statistique.
- Les unités statistiques ou bien les individus correspondent aux étudiants.
- La population est l'ensemble des étudiants.
- Lors de l'étude sur l'emploi la population d'étude pourrait être l'ensemble des personnes en âge de travailler.

2. **Variable ou caractère:** Chaque individu d'une population est décrit par un ensemble de caractéristiques appelées variables ou caractères, on les représente souvent par des lettres majuscules: X, Y,....

Les valeurs qui peut prendre une variable statistique sont appelées modalisées.

Une variable doit donc présenter au minimum deux modalités

- **Exemple2:**

- La variable sexe a deux modalités: masculin ou féminin.
- Les modalités de la variable âge des ouvriers d'une entreprise peuvent être: [25, 30[; [30, 35[; [35, 40[; [40, 45[; [45, 50[et [50, 60[.
- Les variables statistiques peuvent être classées selon leur nature en deux catégories:

1) **Variable quantitative:** une variable est dite quantitative si ses différentes valeurs sont mesurables. Il y a deux types de variables quantitatives:

a. **Variable discrète:**

Une variable quantitative est dite **discrète** si elle prend un nombre fini ou dénombrable de valeurs.

- **Exemple3:** La variable nombre d'enfants par famille, dont les valeurs sont:0, 1, 2, 3,4etc... est une variable quantitative discrète.

b. **Variable continue:** Une variable quantitative est dite **continue** si elle prend toutes les valeurs dans des intervalles de l'ensemble des nombres réels  $\mathbb{R}$ .

- **Exemple 4:** La variable taille en cm observée sur un échantillon de 500 individus est une variable quantitative continue, dont les résultats sont les suivants:

- 100 individus leurs tailles sont comprise entre dans l'intervalle [150, 160[; 250 individus dans l'intervalle [160, 170[; 150 dans l'intervalle [170, 185].

2) **Variable qualitative:** La variable est dite qualitative si les différentes modalités ne sont pas mesurables.

- **Exemple5:** le sexe, la nationalité, l'état matrimonial, la couleur des yeux, la catégorie socioprofessionnelle.

**2.3 Observations:** Une variable donne lieu à plusieurs observations. Si la variable est notée **X**, alors on notera par  $x_1, x_2, x_3, \dots, x_n$  les observations ou série statistique, où **n** présente le nombre d'observations.

Si les observations portent sur la totalité de la population, on dit que **n** est la taille (ou l'effectif) de la population, si les observations n'ont porté que sur une partie de la population, on dit que **n** est la taille de l'échantillon de la population observée.

- **Exemple6:** Soit un échantillon de dix individus, la variable âge observé peut prendre les valeurs suivantes:  $x_1=31, x_2=40, x_3=25, x_4=30, x_5=35, x_6=40, x_7=28, x_8=42, x_9=33, x_{10}=37$ .

Dans ce cas la taille de l'échantillon est  $n=10$ .

**2.4. Données:** Lorsque les observations portent sur une ou plusieurs variables (on note **p** variable) et sur un ou plusieurs individus (on note **n** individus), l'ensemble des valeurs des observations récoltées constituent les données de l'étude ou de l'expérience effectuée. On peut les mettre sous forme d'un tableau individus/variable à **n** lignes et **p** colonnes.

variables \ Individus	X <sub>1</sub>	X <sub>2</sub>	.....	X <sub>p</sub>
l <sub>1</sub>	X <sub>11</sub>	X <sub>12</sub>	.....	X <sub>1p</sub>
l <sub>2</sub>	X <sub>21</sub>	X <sub>n1</sub>	.....	X <sub>2p</sub>
·	·	·	·	·
·	·	·	·	·
·	·	·	·	·
l <sub>n</sub>	X <sub>n1</sub>	X <sub>n2</sub>		X <sub>np</sub>

Chaque individu **I** est représenté par un numéro d'ordre  $i=1, 2, \dots, n$ ; chaque variable **X** est indiquée d'un numéro d'ordre correspondant  $j=1, 2, \dots, p$ .

$x_{ij}$  est la valeur prise par la  $j^{\text{ème}}$  variable pour le  $i^{\text{ème}}$  individu.

**Exemple7:**

Soit un échantillon de dix individus, sur lesquels on observe deux variables X<sub>1</sub> "le poids" et X<sub>2</sub> "l'âge", les résultats de cette étude sont représentés dans le tableau suivant:

variables \ Individus	X <sub>1</sub> = poids	X <sub>2</sub> = âge
l <sub>1</sub>	40	12
l <sub>2</sub>	50	18
l <sub>3</sub>	65	25
l <sub>4</sub>	75	30
l <sub>5</sub>	80	42
l <sub>6</sub>	72	37
l <sub>7</sub>	78	25
l <sub>8</sub>	68	38
l <sub>9</sub>	65	19
l <sub>10</sub>	64	24

## Leçon2:

### Les tableaux statistiques

#### 1. Présentation des données sous forme de tableaux:

Au cours d'une étude les données sont recueillies de façon désordonnée.

Cette masse d'informations que si les informations en question sont classées et mises en ordre, ce classement ne peut se faire que dans des tableaux statistiques qui servent de documentation statistique.

Ces tableaux doivent comporter:

- Le titre indiquant l'objet du tableau.
- L'unité utilisée (Dinars, milliers de tonnes, milliers de personnes, le pourcentage....).
- Les titres des lignes et des colonnes
- Les sources qui doivent être citées.

Leur présentation diffère selon la nature des variables

##### a. Cas d'une variable discrète:

Lorsque la variable statistique est discrète, il arrive souvent qu'une même valeur  $x_i$  ( $i = 1, 2, 3, \dots, k$ ;  $k$  = nombre de modalité) soit observée sur plusieurs individus, on note alors  $n_i$  son effectif (fréquence absolue), qui est le nombre d'observations égales à  $x_i$  et  $f_i$  la fréquence relative correspondante:  $f_i = n_i/n$ ,  $i = 1, 2, \dots, k$ .

Le tableau statistique se présente alors sous la forme suivante:

$x_i$	$n_i$	$f_i$
-------	-------	-------

##### - Remarque:

$$\sum_{i=1}^k n_i = \sum_{i=1}^k f_i = 1$$

##### - Exemple1:

Distribution statistique du personnel d'une entreprise d'après le nombre d'enfants à charge:

Nombre d'enfants à charge	Effectifs $n_i$
0	4
1	15
2	29
3	18
4	10
5	3
6	1
Total	<b>80 = somme des <math>n_i = \sum_{i=1}^{i=7} n_i = N</math></b>

Dans cet exemple la variable discrète prend sept valeurs:

$$x_1=0 \quad x_2=1 \quad x_3=2 \quad x_4=3 \quad x_5=4 \quad x_6=5 \quad x_7=6$$

A chacune des ces valeurs correspond son effectif:

$$n_1=4 \quad n_2=15 \quad n_3=29 \quad n_4=18 \quad n_5=10 \quad n_6=3 \quad n_7=1$$

La série statistique (ou distribution statistique) est l'ensemble des sept couples:

$$(x_1, n_1) \quad (x_2, n_2) \quad (x_3, n_3) \quad (x_4, n_4) \quad (x_5, n_5) \quad (x_6, n_6) \quad (x_7, n_7)$$

Soit:

$$(0, 4) \quad (1, 15) \quad (2, 29) \quad (3, 18) \quad (4, 10) \quad (5, 3) \quad (6, 1)$$

Et nous dirons la série  $(x_i, n_i)$ ,  $i$  (indice  $i$  prenant les valeurs: 1, 2, 3, 4, 5, 6, 7).

La somme des effectifs est l'effectif total  $N$  de la population étudiée:

$$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7$$

Que nous noterons:

$$\sum_{i=1}^{i=7} n_i$$

Souvent nous nous contentons d'écrire  $\sum n_i$  (qu'on lira somme des  $n_i$ ) pour noter la somme de tous les effectifs de la distribution statistique étudiée.

**b. Cas d'une variable continue:**

Lorsque la variable est continue, on regroupe les observations en  $k$  classes d'extrémités:  $c_0, c_1, c_2, c_3, \dots, c_k$

On note pour chaque classe  $[c_i, c_{i+1}[$  l'effectif  $n_i$  et la fréquence relative  $f_i$ , le tableau statistique se présente sous la forme suivante:

$c_i$	$n_i$	$f_i$
$c_{i+1}$		

Ou bien

$[c_i, c_{i+1}[$	$n_i$	$f_i$
------------------	-------	-------

Par convention la borne supérieure d'une classe est exclue de cette classe, sauf pour la dernière classe (dernière classe peut être fermée).

On appelle centre d'une classe  $[c_i, c_{i+1}[$  la valeur:

$$y_i = \frac{c_i + c_{i+1}}{2} \quad \text{et l'amplitude de cette est notée } a_i = c_{i+1} - c_i$$

### Exemple2:

On désire présenter la répartition de la population algérienne résidente de 2008

Tranche d'âge en années	Population en 2008
0-1	823000
1-4	1245700
4-9	1345200
9-13	1457000
.	.
.	.
.	.
75-80	225000
80 et plus	122500
total	23455000

### Sources ONS

Ainsi la lecture est facile et le regroupement peut se faire aussi selon une autre amplitude.

**Exemple3:** On a observé les tailles en cm de 24 personnes, les résultats sont les suivants: 170, 165, 170, 165, 175, 165, 178, 178, 162; 159, 158, 176, 180, 180, 182, 179, 183, 165, 176, 177, 180, 173, 173, 162.

Ces observations peuvent être représentées par des classes d'amplitude constantes égale à 4 cm par exemple:

Tailles	Effectif partiel $n_i$	Fréquences $f_i$	Fréquences en pourcentage
[158 – 162[	2	0.083	8.3%
[162 – 166[	6	0.25	25%
[166 – 170[	0	0	0%
[170 – 174[	4	0.167	16.7%
[174 – 178[	4	0.167	16.7%
[178 – 182[	6	0.25	25%
[182 – 186[	2	0.083	8.3%
Total	24	1	100%

**Exemple 4:** Une enquête sur la destination de 12 touristes algériens a permis d'avoir les résultats suivants:

Touriste N°	Destination
1	Tunisie
2	Turquie
3	Turquie
4	Tunisie
5	Tunisie
6	France
7	Espagne
8	France
9	Turquie
10	Tunisie
11	Maroc
12	Turquie

La variable étudiée représente **les destinations des touristes algériens**, elle présente **cinq modalités**: Tunisie, France, Turquie, Maroc, Espagne avec les **effectifs partiels** respectivement  $n_1=4$ ,  $n_2=3$ ,  $n_3=3$ ,  $n_4=1$ ,  $n_5=1$ . De là, on obtient le tableau statistique correspondant:

Destination	Effectifs partiels $n_i$	Fréquences $f_i$
Tunisie	4	0.333
France	3	0.25
Turquie	3	0.25
Maroc	1	0.083
Espagne	1	0.083
Total	12	1

**Remarque:**

*Dans l'exemple précédent, la somme des fréquences n'est pas égale à 1, ceci est dû aux calculs effectués sur les fréquences. En effet on ne peut pas prendre tous les chiffres après la virgule, par contre on peut arrondir certain chiffre de manière à avoir la somme égale à 1.*

2. **Règle de STURGES:** Cette règle est utilisée pour déterminer le nombre de classes à utiliser pour représenter une variable statistique continue.

Le nombre de classes est égal à l'entier le plus proche de la quantité:

$$1 + 3.3 \log n_i$$

Par exemple pour un échantillon de taille  $n = 200$ , on doit utiliser:

$$1 + 3.3 \log 200 = 1 + 3.3 \times 2.3 = 8.59 = 9 \text{ classes}$$

L'amplitude constante de ces classes sera égale à:

$$a = \frac{X_{\max} - X_{\min}}{\text{nombre de classes}}$$

3. **Distribution et les effectifs cumulés:**

3.1. **Distribution:** La distribution statistique X est l'ensemble des couples  $(x_i, f_i)$ ,  $i = 1, \dots, k$  (k nombre de modalités de la variable étudiée).

**Exemple5:** (0; 0.2), (1; 0.3), (2; 0.4), (3 et +; 0.1) est une distribution statistique représentant la variable nombre d'enfants par famille.

3.2. **Fonction de répartition:** La fonction de répartition appelée aussi fonction des fréquences cumulées associée à tout nombre réel x la proportion des individus pour lesquels la variable statistique X prend des valeurs inférieure à x, donc  $F(x) = \sum f_i$  exemple6:

**Exemple 6:** Le taux de glycémie, observé sur un échantillon de 100 patients sont donnés dans le tableau suivants:

Taux de glycémie $x_i$	[0.8-0,85 [	[0.85-0,9 [	[0.9-0,95[	[0,95-1[	[1-1,20 [	total
Effectif $n_i$	10	20	30	30	10	100
Fréquences $f_i$	0,1	0,2	0,3	0,3	0,1	1
F (x)	0	0,1	0,3	0,6	0,9	1

Pour:

$$x = 0,8; F(0,8) = \sum_{x_i < 0,8} f_i = 0$$

$$x = 0,85; F(0,85) = \sum_{x_i < 0,85} f_i = f_1 = 0,1$$

$$x = 0,9; F(0,9) = \sum_{x_i < 0,9} f_i = f_1 + f_2 = 0,3$$

$$x = 0,95; F(0,95) = \sum_{x_i < 0,95} f_i = f_1 + f_2 + f_3 = 0,6$$

$$x = 1; F(1) = \sum_{x_i < 1} f_i = f_1 + f_2 + f_3 + f_4 = 0,9$$

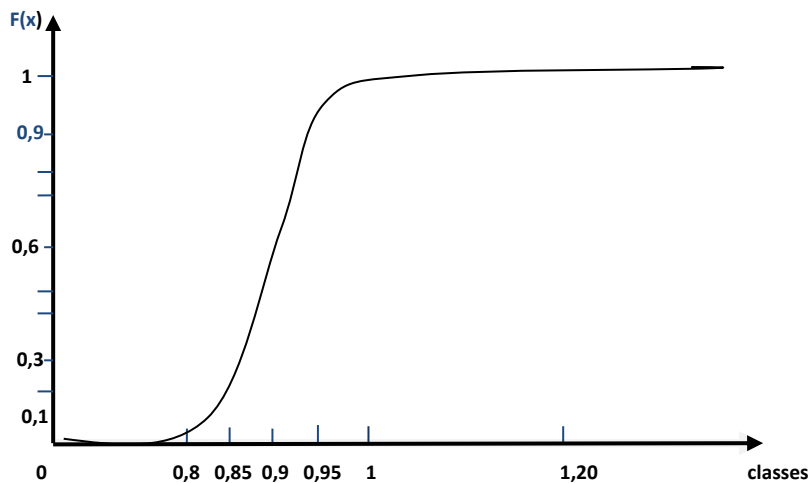
$$x = 1,20; F(1,20) = \sum_{x_i < 1,20} f_i = f_1 + f_2 + f_3 + f_4 + f_5 = 1$$

$F(0,95)=0,6$  par exemple, veut dire que 60% des patients ont un taux de glycémie inférieure à 0,95.

**Remarque:**

- On a  $F(x)=0$ , si  $x \leq \min x_i$  et  $F(x) = 1$ , si  $x > \max x_i$
- $F$  est une fonction continue croissante, elle prend ses valeurs dans l'intervalle  $[0, 1]$  et présente lorsque  $X$  est discrète une discontinuité en chacun des points  $x_i$  qui se manifeste par un saut dont la valeur est  $f_i$ .

Ainsi la fonction de répartition précédente peut être représentée graphiquement par une courbe continue (courbe cumulative).

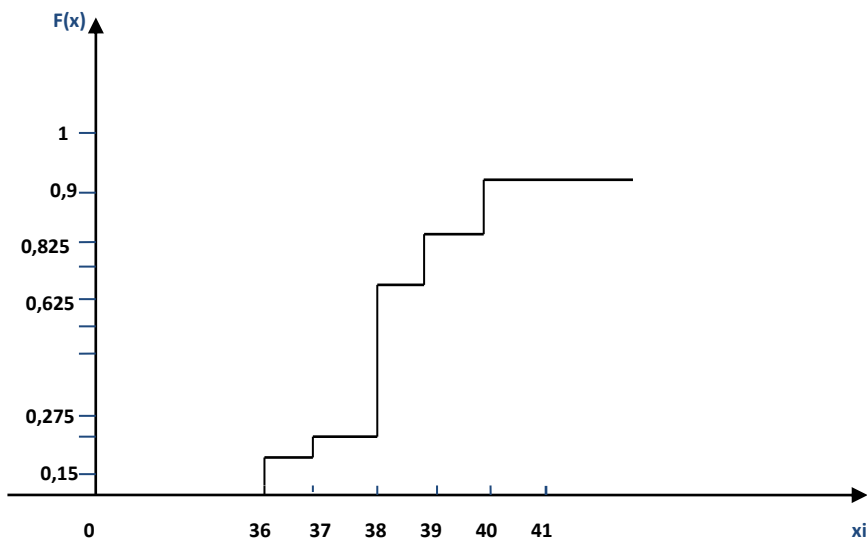


**Exemple 7:** une enquête sur les pointures de 200 filles a fourni les résultats suivants:

$x_i$	36	37	38	39	40	41	Total
$n_i$	30	25	70	40	15	20	200
$f_i$	0.15	0.125	0.35	0.2	0.075	0.1	1
$F(x)$	0	0.15	0.275	0.625	0.825	0.9	1

Comme la variable étudiée est discrète, alors la fonction de répartition va être représentée par une courbe en escalier.





### 3.2 Courbe cumulative croissante et décroissante:

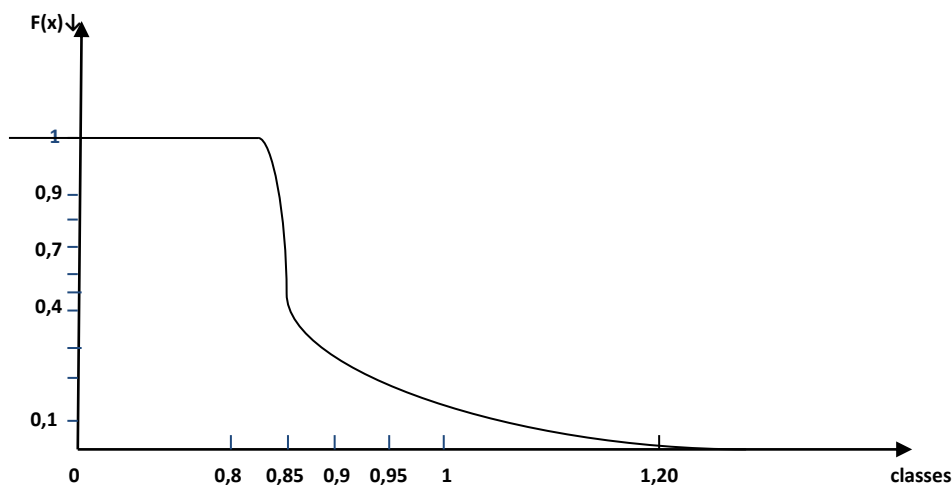
Pour définir la fonction de répartition ; on calcule les fréquences cumulées croissantes qu'on représente graphiquement par une courbe cumulative croissante.

De la même façon on peut calculer aussi les fréquences cumulées décroissantes, à qui correspond une courbe cumulative décroissante, notée :

$$F(x) \downarrow \sum_{x_i \geq X} f_i$$

Calculons les fréquences cumulées décroissantes de la variable étudiée dans l'exemple 6

Taux de glycémie $x_i$	[0,8-0,85 [	[0,85-0,9 [	[0,9-0,95[	[0,95-1[	[1-1,20 [	total
Effectif $n_i$	10	20	30	30	10	100
Fréquences $f_i$	0,1	0,2	0,3	0,3	0,1	1
$F(x) \downarrow$	1	0,9	0,7	0,4	0,1	0
$F(x) \uparrow$	0	0,1	0,3	0,6	0,9	1

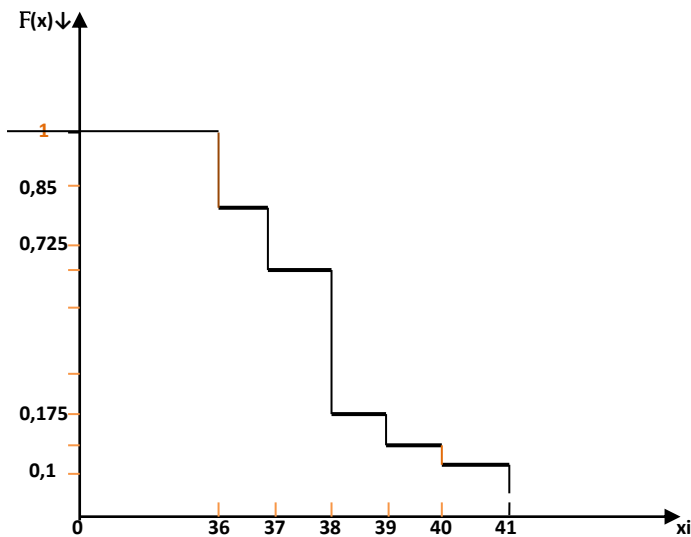


$F(0,9) \downarrow$ , c'est-à-dire 70% des patients ont un taux de glycémie  $\geq 0,9$ .

$F(0,9) \downarrow$ , c'est-à-dire 30% des patients ont un taux de glycémie  $< 0,9$ .

Revenons à l'exemple 7

$x_i$	36	37	38	39	40	41	Total
$n_i$	30	25	70	40	15	20	200
$f_i$	0.15	0.125	0.35	0.2	0.075	0.1	1
$F(x) \downarrow$	1	0.85	0.725	0.375	0.175	0.1	0



**Remarque :** les courbes cumulatives croissantes et décroissantes peuvent être obtenues aussi à l'aide des effectifs cumulés croissant (notés  $N_i \uparrow$ ) et les effectifs cumulés décroissant (notés  $N_i \downarrow$ ).

## Les représentations graphiques

### 1. Introduction :

Bien qu'un tableau statistique contient toute l'information prélevé sur l'échantillon mais il est très utile de traduire ces informations par un graphique, afin d'avoir une synthèse visuelle.

Suivant la nature de la variable étudiée, on utilise différents types de représentation.

**2. Cas de variable qualitative:** Une variable qualitative peut être représentée graphiquement de trois manières.

**2.1. Représentation circulaire par des secteurs:** Elle consiste à représenter sur un cercle chaque modalité par un secteur dont le degré est déterminé à l'aide de la règle de trois de la manière suivante:

Supposons qu'on veut calculer le degré du secteur représentant la modalité  $i$  d'effectif partiel  $n_i$  alors on peut écrire :

$$n \text{ (l'effectif total)} \rightarrow 360^\circ$$

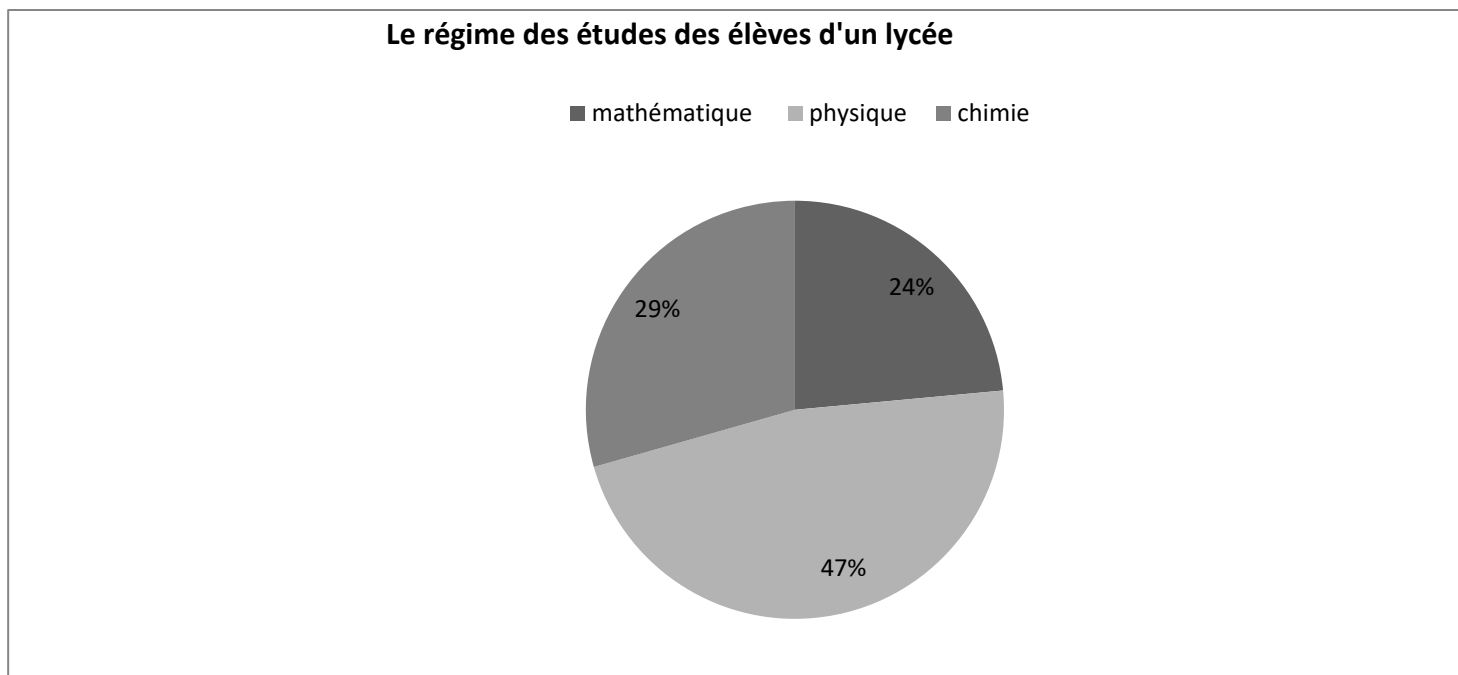
$$\rightarrow d_i = \frac{n_i \times 360}{n} = f_i \times 360$$

**Exemple1** : On étudié le régime des études sur un échantillon de 200 élèves d'un lycée donné, les résultats obtenus sont les suivants :

Régime des études	Externe	Interne	Demi-pensionnaire
Nombre d'élèves $n_i$	70	50	80
Fréquences $f_i$	0.35	0.25	0.4

Calculons le degré correspondant aux différents secteurs :

- $d_i = f_i \times 360^\circ = 0.35 \times 360^\circ = 126^\circ$
- $d_i = f_i \times 360^\circ = 0.25 \times 360^\circ = 90^\circ$
- $d_i = f_i \times 360^\circ = 0.4 \times 360^\circ = 144^\circ$



## 2.2. Représentation en tuyau d'orgue :

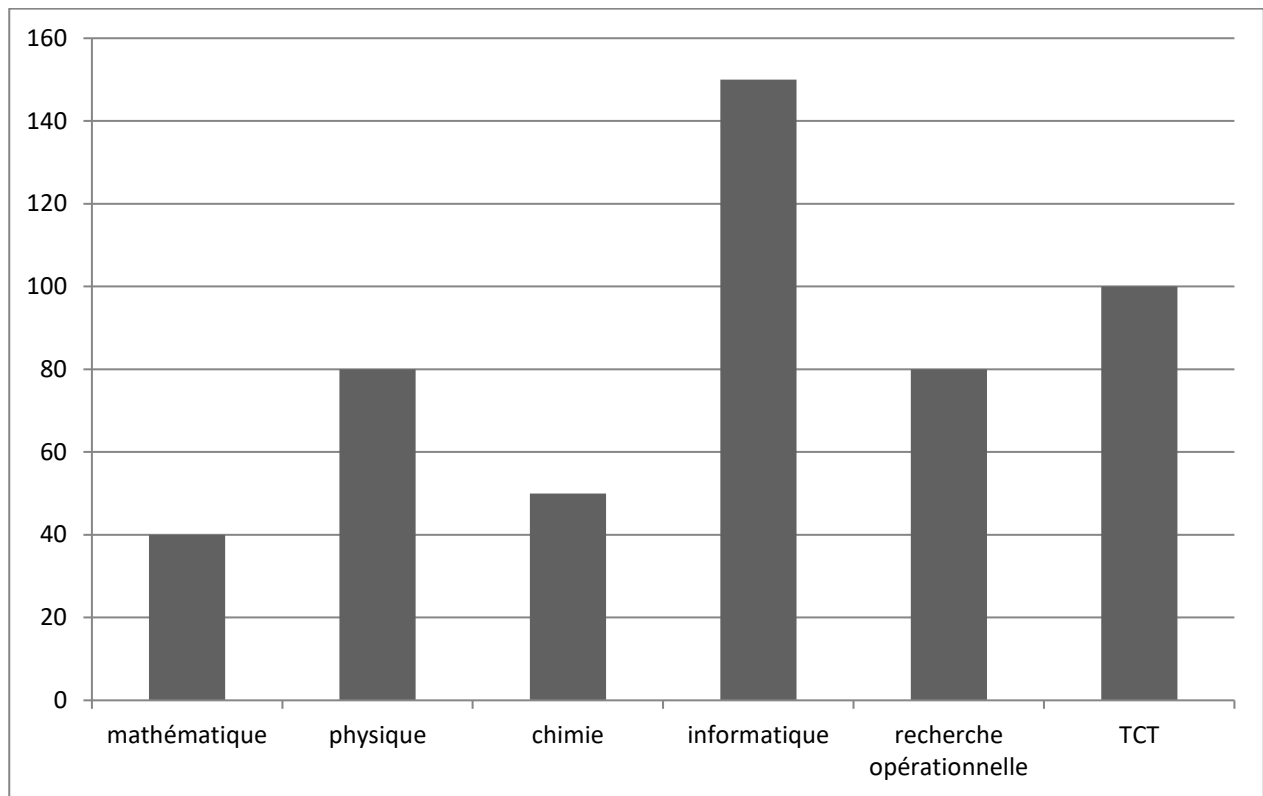
Elle consiste à porter en abscisse les modalités de la variable étudiée, au dessus des quelles on trace des tuyaux d'orgues de même largeur, et en ordonnée les effectifs ou les fréquences qui vont déterminer leurs hauteurs.

### **Exemple2 :**

Une enquête auprès de 500 étudiants de 1<sup>ère</sup> année SETI (sciences exactes, technologie et informatique) admis en 2<sup>ème</sup> année sur les spécialités préférées a fourni les résultats suivants :

spécialité préférée	nombre d'étudiants
mathématique	40
physique	80
chimie	50
informatique	150
recherche opérationnelle	80
TCT	100
Total	500

Ces résultats peuvent être traduits par un diagramme de tuyaux d'orgue suivant



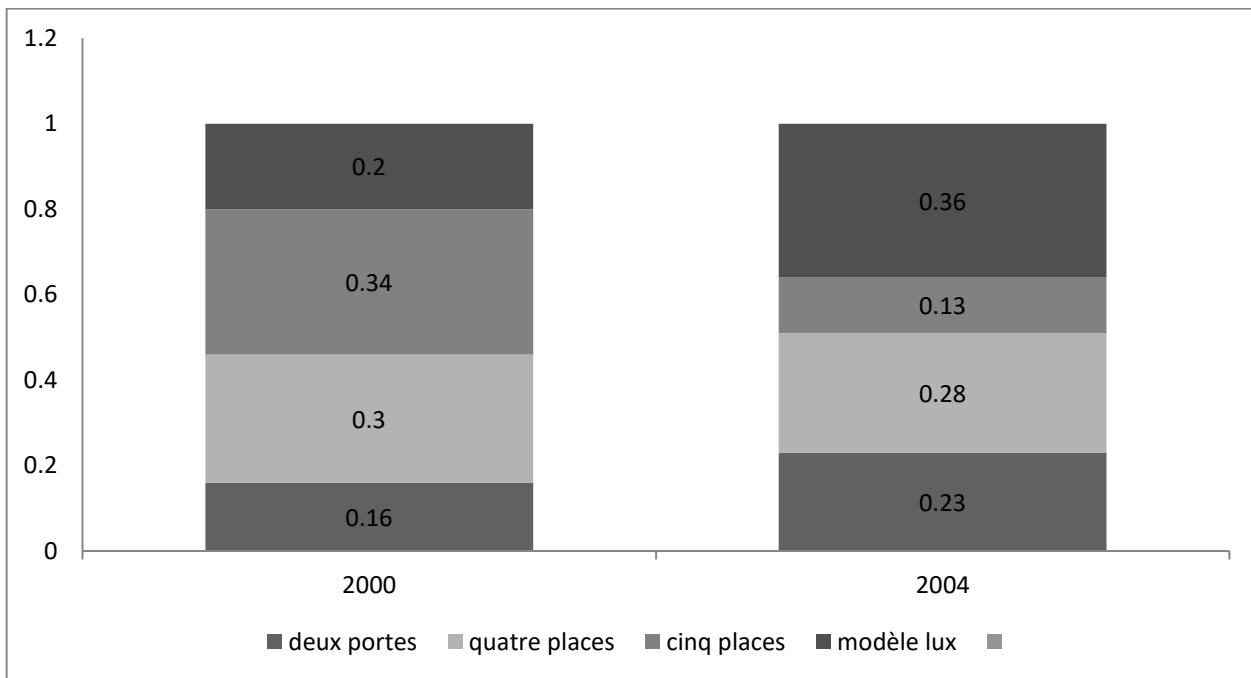
### 2.3. Diagramme en bandes:

Il consiste à représenter dans une même bande verticale, chaque modalité par une tranche dont la hauteur correspond à sa fréquence.

#### Exemple3:

Les ventes réalisées par une maison de fabrication d'automobiles durant les années 2000 et 2004 sont les suivantes:

véhicules	2000		2004	
	$n_i$	$f_i$	$n_i$	$f_i$
Deux portes	800	0.16	1600	0.23
Petites 4 places	1500	0.3	2000	0.28
Cinq places	1700	0.34	900	0.13
Modèle luxe	1000	0.2	2500	0.36
Total	5000	1	7000	1



### 3. Cas de variable quantitative:

On va distinguer les diagrammes correspondant à une variable discrète et ceux correspondant à une variable continue.

#### 3.1. Le diagramme en bâton :

Il consiste à représenter les divers valeurs  $x_i$  prise par une variable discrète, en fonction des fréquences  $f_i$  ou les effectifs  $n_i$  de la manière suivante :

On représente en abscisse les observations  $x_i$  et au dessus de chaque valeur en trace un bâton dont la hauteur et proportionnelle à  $f_i$  ou  $n_i$ .

#### Exemple4 :

Les poids (en kg) de 200 individus sont représentés dans le tableau suivant :

$x_i$	55	60	65	70	75	80	Total
$n_i$	20	60	40	20	50	10	200
$f_i$	0.1	0.3	0.2	0.1	0.25	0.05	1

Cette variable est discrète, donc on peut la représenter graphiquement par le diagramme en bâton en utilisant les effectifs  $n_i$  ou les fréquences  $f_i$ .

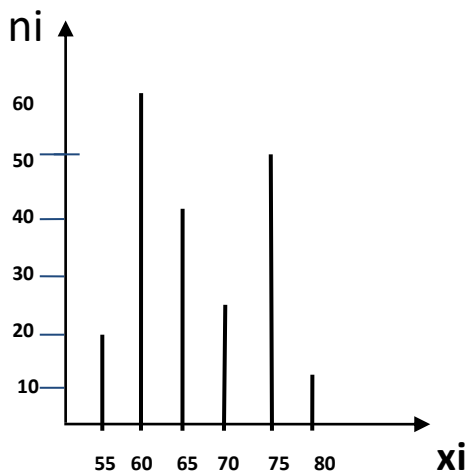


Diagramme en bâton

### 3.2. Histogramme:

L'histogramme est utilisé pour représenter une variable continue.

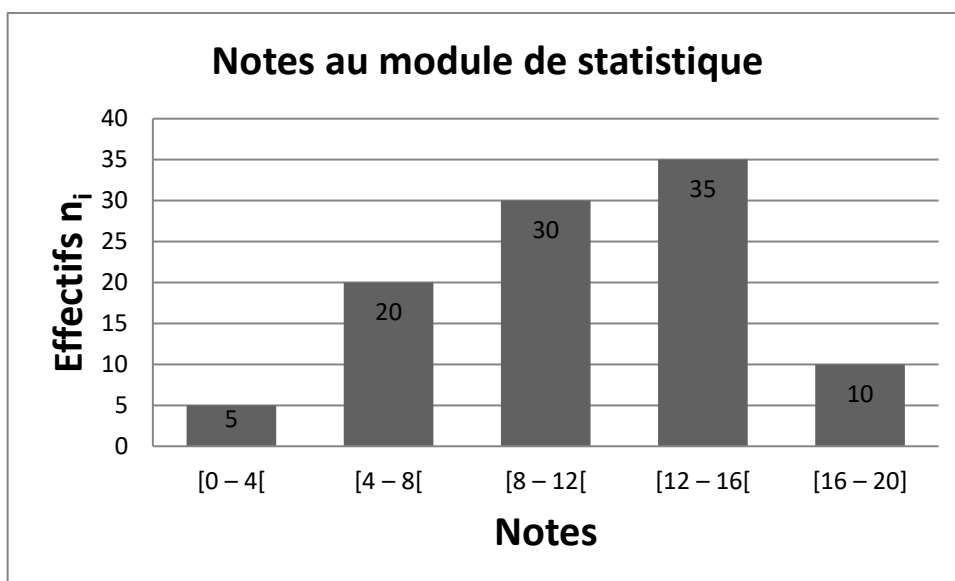
On représente sur l'axe des abscisses les différentes classes supposée de même amplitude et au dessus de chacune d'elle on trace des rectangles, dont les hauteurs sont discrètement proportionnelles aux fréquences ou aux effectifs.

#### Exemple5 :

Les notes au module de statistique observées sur un échantillon de 100 étudiants sont données dans le tableau suivant:

Notes	[0 – 4[	[4 – 8[	[8 – 12[	[12 – 16[	[16 – 20]	Total
Effectifs $n_i$	5	20	30	35	10	100

La variable étudiée est continue, donc elle peut être représentée graphiquement au moyen de l'histogramme suivant :



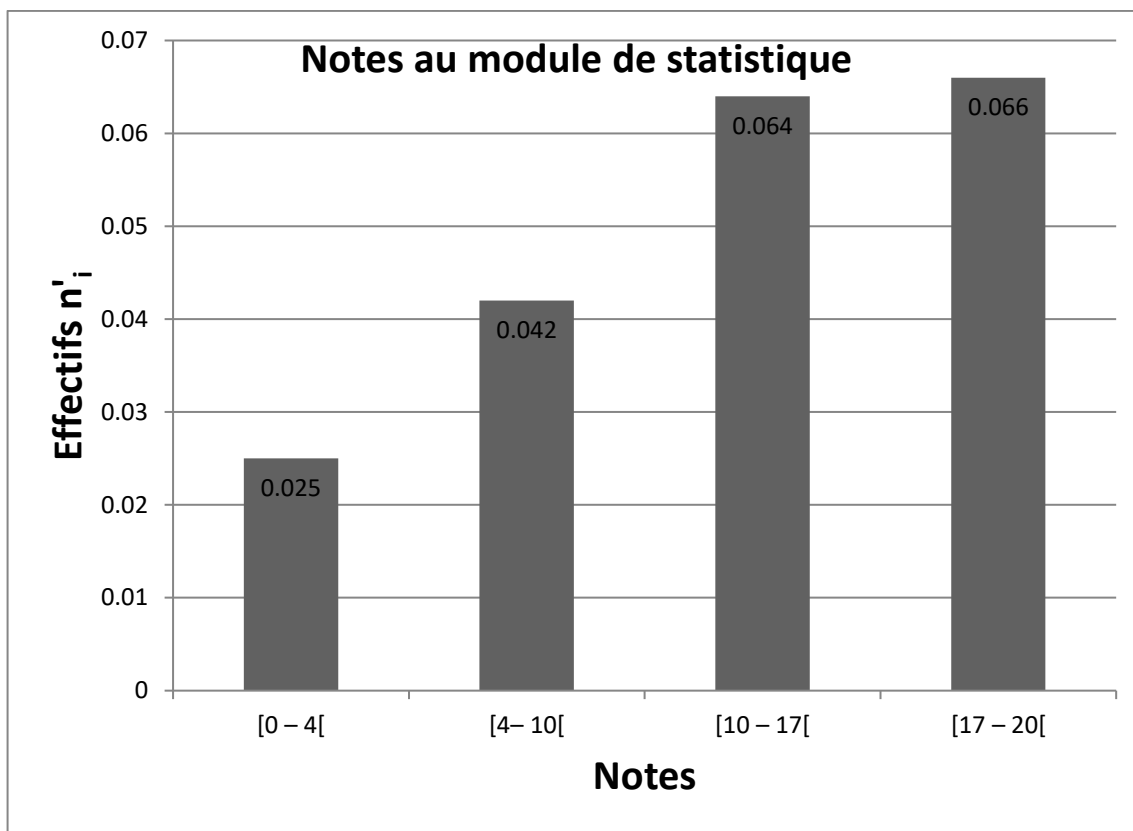
Remarque: ce mode de représentation est valable seulement dans le cas où on a des classes d'intervalles d'amplitude constante.

Dans le cas où les amplitudes sont différentes, pour tracer l'histogramme on utilise la méthode des effectifs rectifiés, qui consiste à multiplier l'effectif de chaque classe par le rapport entre l'amplitude de la plus petite classe sur l'amplitude de la classe considérée, on obtient ainsi les effectifs rectifiés noté  $n_i^{\circ}$  :

$$n_i^{\circ} = \frac{\text{la plus petite amplitude}}{\text{amplitude de la classe } i}$$

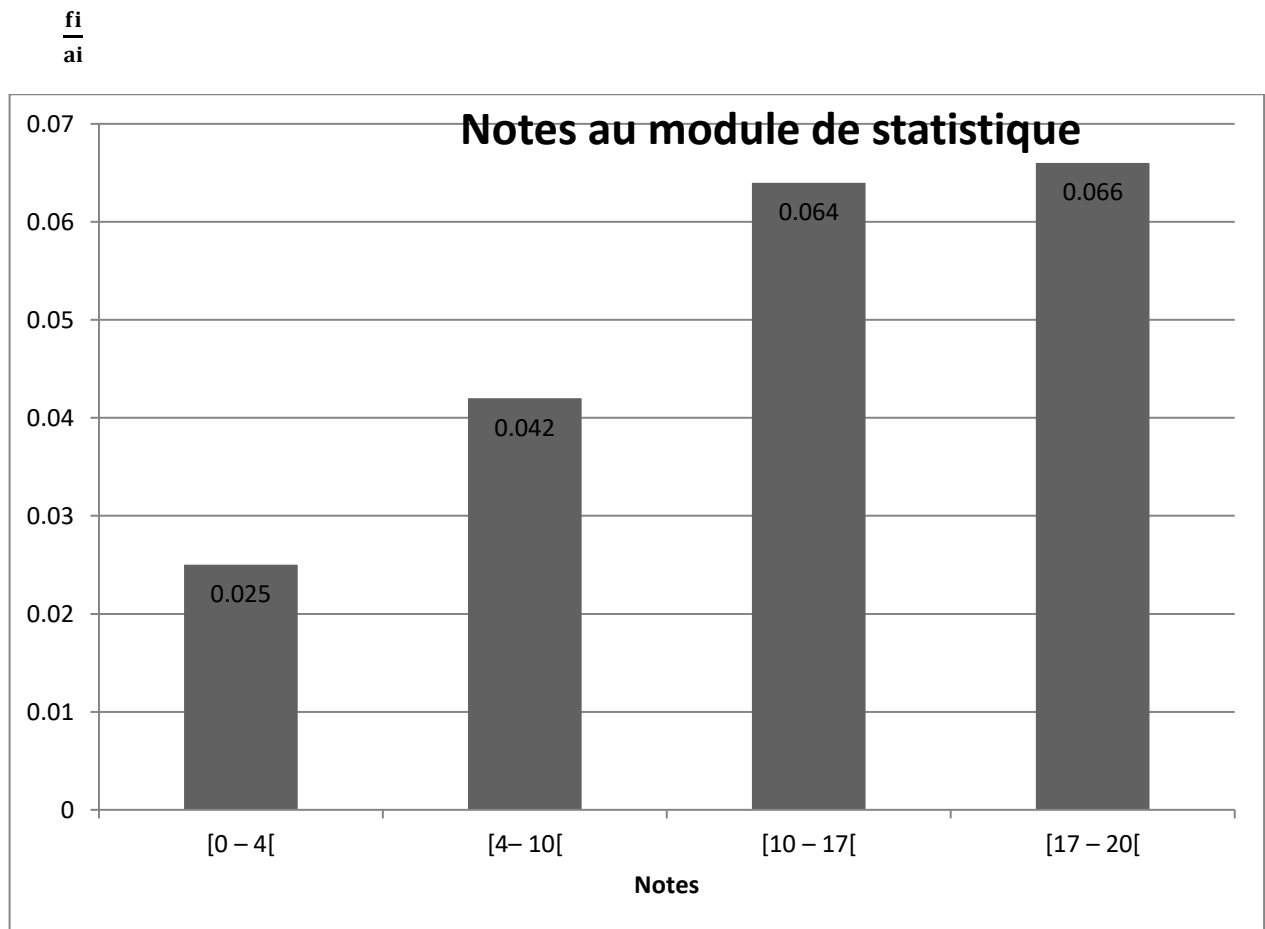
**Exemple6 :**

Notes	[0 – 4[	[4 – 10[	[10 – 17[	[17 – 20]	Total
Effectifs $n_i$	10	25	45	20	100
Amplitude $a_i$	4	6	7	3	/
Effectifs rectifié $n_i^{\circ}$	$10 \times \frac{3}{4} = 7$	$25 \times \frac{3}{6} = 13$	$45 \times \frac{3}{7} = 19$	$20 \times \frac{3}{3} = 20$	/



On peut aussi représenter l'histogramme autrement, en utilisant comme hauteurs des rectangles les quantités :  $\frac{f_i}{a_i}$  ou  $(\frac{n_i}{a_i})$ .

Notes	[0 – 4[	[4 – 10[	[10 – 17[	[17 – 20]	Total
Effectifs $n_i$	10	25	45	20	100
Fréquences $f_i$	0.1	0.25	0.45	0.2	1
Amplitude $a_i$	4	6	7	3	/
Fréquences rectifié $\frac{f_i}{a_i}$	$\frac{0.1}{4} = 0.025$	$\frac{0.25}{6} = \mathbf{0.042}$	$\frac{3}{7} = \mathbf{0.064}$	$\frac{0.2}{3} = \mathbf{0.066}$	/





- Les moyennes :

1- La moyenne arithmétique

La moyenne arithmétique d'une variable statistique est égale à la somme des valeurs prise par cette variable, divisée par le nombre d'observations.

On la note par  $\bar{x}$  ou  $\mu$  telle que  $\bar{x} = \frac{1}{n} \sum x_i$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum x_i$$

Exemple : la moyenne arithmétique des nombres : 11. 10 5. 12. 13. 8. 7 est :

$$\bar{x} = \frac{11+10+5+12+13+8+7}{7} = \frac{66}{7} = 9.43$$

La moyenne arithmétique est dite simple car chaque valeur correspond à un individu et par opposition la moyenne arithmétique pondérée.

- La moyenne arithmétique pondérée :

Soit une variable statistique dont les différentes modalités :  $x_1, x_2, x_3, \dots, x_k$  auxquelles correspondent respectivement les effectifs :  $n_1, n_2, n_3, \dots, n_k$  avec :  $n_1 + n_2 + n_3 + \dots + n_k = n$

la moyenne arithmétique pondérée de telle variable est :

$$\bar{x} = \frac{(x_1 \times n_1) + (x_2 \times n_2) + (x_3 \times n_3) + \dots + (x_k \times n_k)}{n} = \frac{1}{n} \sum x_i \times n_i$$

Ou encore :  $\bar{x} = \sum \frac{n_i}{n} \times x_i = \sum f_i \times x_i$

Remarque :

Si  $x$  est une variable continue alors  $\bar{x} = \frac{1}{n} \sum n_i \times c_i$ , avec  $c_i$  centre de classe.

Exemple :

poids en kg $x_i$	nombre d'individus $n_i$	$c_i$	$n_i \times c_i$
[56-58[	5	57	285
[58-60[	36	59	2124
[60-62[	40	61	2440
[62-64[	15	63	945
[64-66[	4	65	260
total	100		6054

$$\bar{x} = \frac{1}{n} \sum n_i \times c_i = \frac{1}{100} (5 \times 57) + (36 \times 59) + (40 \times 61) + (15 \times 63) + (4 \times 65) = 60,54$$

## 2- La moyenne géométrique:

La moyenne géométrique simple G d'une série de n valeurs :  $x_1, x_2, x_3, \dots, x_n$  est la racine n<sup>ième</sup> du produit de ces valeurs.

### a. Cas d'une variable discrète :

$$G = \sqrt[n]{x_1 \times x_2 \times x_3 \dots \times x_n}$$

En prenant le logarithme des deux membres on obtient :

$$\log G = \frac{1}{n} \sum \log x_i$$

#### - Exemple :

La moyenne géométrique des valeurs 2, 3, 8, 11, 17 est :  $G = \sqrt[5]{2 \times 3 \times 8 \times 11 \times 17} = 6.17$

### b. Cas d'une variable continue :

Partons du tableau statistique suivant :

$x_i$	$n_i$
1.08	3
1.07	2
1.09	5
Total	10

Et calculons la moyenne arithmétique prise par la variable x

$x_i$	$n_i$	$x_i n_i$
1,08	3	3,24
1,07	2	2,14
1,09	5	5,45
Total	10	10,83

$$m = \frac{\sum x_i n_i}{\sum n_i} = \frac{10.83}{10} = 1.083$$

$x_i$	$n_i$	$\log x_i$	$n_i \log x_i$
1,08	3	0,03342	0,10026
1,07	2	0,02938	0,05876
1,09	5	0,03743	0,18715
Total	10		0,34617

$$\log G = \frac{\sum n_i \times \log x_i}{\sum n_i} = \frac{0,34617}{10} = 0,034617$$

- D'où  $G = \mathbf{1,082975}$

### 3 - La moyenne harmonique :

La moyenne harmonique simple H d'une série de n valeurs  $x_1, x_2, \dots, x_n$  est l'inverse de la moyenne arithmétique des inverses des valeurs :

$$H = \frac{1}{\frac{1}{n} \sum \frac{1}{x_i}} = \frac{n}{\sum \frac{1}{x_i}}$$

Si on a une variable statistique ayant k valeurs  $x_1, x_2, \dots, x_k$  aux quelles correspondent respectivement les effectifs :  $x_1, x_2, x_3, \dots, x_n$  alors la moyenne harmonique pondérée de ces valeurs est :

$$H = \frac{1}{\frac{1}{n} \sum \frac{n_i}{x_i}} = \frac{n}{\sum \frac{n_i}{x_i}}$$

#### Exemple 1 :

La moyenne harmonique des valeurs 1, 4 ; 8 ; 10 ; 12 ; est

$$H = \frac{5}{\frac{1}{1} + \frac{1}{4} + \frac{1}{8} + \frac{1}{10} + \frac{1}{12}} = \frac{5}{\sum \frac{1}{x_i}} =$$

$x_i$	1	4	8	10	12	Total
$\frac{1}{x_i}$	1	0,25	0,125	0,1	0,083	1,55

$$H = \frac{5}{1.55} = 3,2$$

#### Exemple 2 :

$x_i$	$n_i$	$x_i \times n_i$	$\frac{x_i}{n_i}$
4	800000	3200000	200000
9	450000	4050000	50000
3	3000000	9000000	1000000
5	750000	3750000	150000
total	5000000		1400000

$$H = \frac{\sum n_i}{\sum \frac{n_i}{x_i}} = \frac{5000000}{1400000} = 3,57 \text{ habitants}, \frac{1}{H} = \frac{1}{3,57} = 0,28$$