

# Statistique Descriptive



**Ecole Supérieur de Technologie de Laayoune**

Université Ibn Zohr Agadir

# SOMMAIRE

**Chapitre 1:** Introduction à la statistique

**Chapitre 2:** Représentation des données

**Chapitre 3:** Caractéristiques de tendance centrale

**Chapitre 4:** Caractéristiques de dispersion et de concentration

**Chapitre 5:** Séries doubles

**Chapitre 6:** Indices Statistiques

**Chapitre 7:** Séries Chronologies

# Chapitre 1

## Introduction à la statistique

# I. Définitions

## 1. La statistique

C'est une science qui consiste à collecter les données, les traiter, les analyser, les interpréter et les présenter en vue d'étudier un phénomène donné.

**Exemple:** Statistique descriptive, statistique décisionnelle,....

## 2. Les statistiques

Ce sont les informations et les données collectées auprès des établissements.

**Exemple:** indices de la bourse, taux de chômage, ....

## 3. La statistique descriptive

C'est l'ensemble des méthodes permettant de décrire, de résumer, de présenter les données.

## II. Vocabulaires de base

### 1. Population

C'est l'ensemble des éléments ayant un caractère commun à étudier. On la note  $\Omega$ .

**Exemple** : Les étudiants de la faculté Ain Sebâa.

### 2. Individu ou Unité statistique

C'est l'élément de la population. On le note  $\omega$ .

**Exemple** : étudiant de la faculté Ain Sebâa.

### 3. Echantillon

C'est un sous-ensemble ou une partie de la population. On le note  $E$ .

**Exemple** : les étudiants S1 de la faculté Ain Sebâa.

## II. Vocabulaires de base

### 4. Taille de la population

C'est le nombre d'éléments de la population. On le note  $N$ .

**Exemple** : le nombre d'étudiants de la faculté Ain Sebâa.

### 5. Caractère ou variable statistique

C'est une caractéristique ou propriété qui décrit un individu d'une population.

Elle est observée ou mesurée sur les individus d'une population. On le note  $X$ .

**Exemple**: Notes des étudiants

Il existe deux types de variables statistiques: **Les variables quantitatives**

et **les variables qualitatives**.

## II. Vocabulaires de base

### 5.1 Variable qualitative

C'est une variable dont les modalités ne peuvent pas être mesurées.

**Exemple:** Langues (Arabe, Anglais, Français,...)

On peut distinguer deux types de variables qualitatives:

#### 5.1.1 Variable qualitative nominale

C'est une variable dont les modalités ne sont pas ordonnées.

**Exemple:** Catégorie socioprofessionnelle (Ingénieur ,Technicien , ...)

#### 5.1.2 Variable qualitative ordinale

C'est une variable dont les modalités sont ordonnées.

**Exemple:** Mention (Très bien, Bien, Assez Bien,...)

## II. Vocabulaires de base

### 5.2 Variable quantitative

C'est une variable dont les modalités prennent des valeurs numériques.

**Exemple**: le salaire mensuel d'un fonctionnaire,

On peut distinguer deux types de variables quantitatives:

#### 5.2.1 Variable quantitative discrète

C'est une variable dont les modalités font partie d'un ensemble fini .

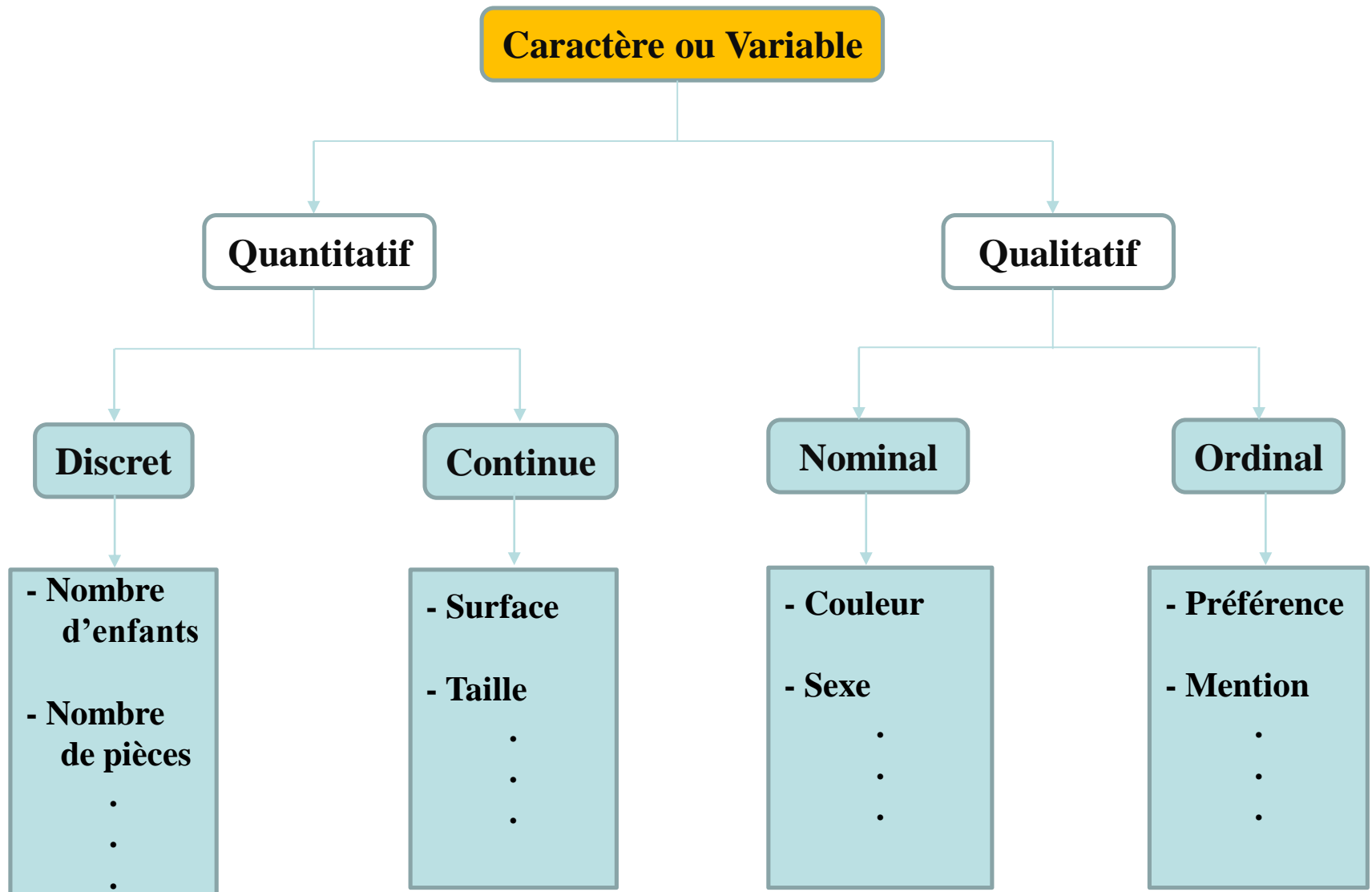
**Exemple**: Le nombre d'enseignants de ESTL.

#### 5.2.2 Variable quantitative continue

C'est une variable dont les modalités prennent des valeurs dans un intervalle réel.

**Exemple**: Poids des cartables des élèves

## II. Vocabulaires de base



## II. Vocabulaires de base

### Exemple 1: Caractère Qualitatif (Nominal)

$\Omega$	= Population	Voitures du parking
$X$	= Caractère	Couleur de voiture
$x_i$	= Modalités	Bleu, Vert, Noire, ...

### Exemple 2: Caractère Qualitatif (Ordinal)

$\Omega$	= Population	Amphi
$i$	= Individu	Etudiant
$X$	= Caractère	Mention
$x_i$	= Modalités	Passable, Assez Bien, Bien, ...

## II. Vocabulaires de base

### Exemple 3: Caractère Quantitatif (Discret)

$\Omega$ = Population	Famille
$X$ = Caractère	nombre d'enfant
$x_i$ = Modalités	0, 2, 3, ...

### Exemple 2: Caractère Quantitatif (Continue)

$\Omega$ = Population	Amphi
$i$ = Individu	Etudiant
$X$ = Caractère	Note
$x_i$ = Modalités	[0,20]

## II. Vocabulaires de base

### 7. Modalité

C'est la valeur d'une variable statistique. On la note  $x$ .

**Exemple:** Les modalités de la variable statistique «  $X = \text{Note}$  » sont : 9,15,11,....

### **Remarque importante:**

Soient une population  $\Omega$  et  $F$  un ensemble des modalités d'un caractère  $X$ .

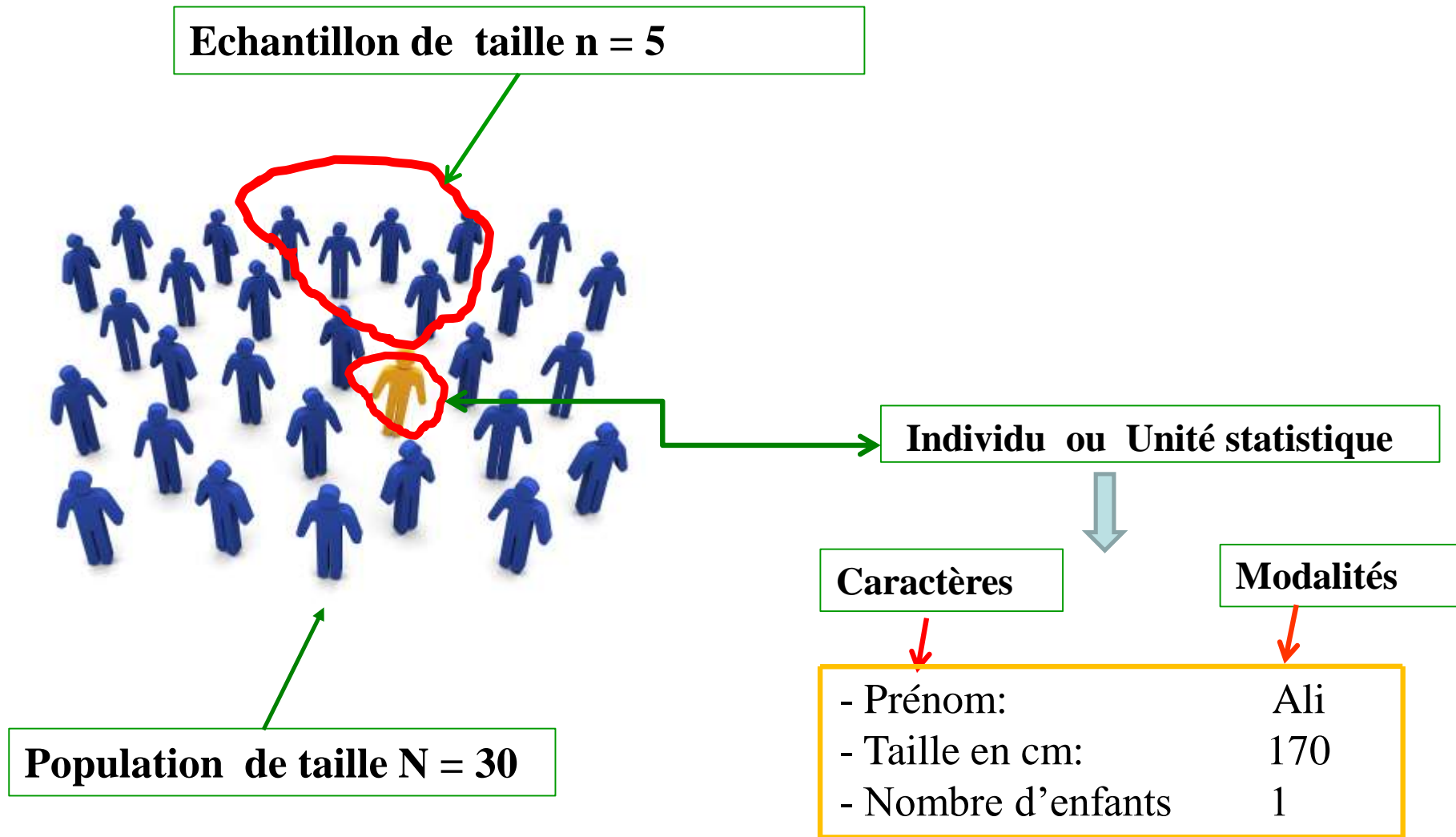
$X$  est une application de  $\Omega$  dans  $F$ , i.e:  $X : \Omega \mapsto F$

$$i \rightarrow X(i) = x_i$$

où :

- $i$  est un individu de  $\Omega$
- $X(i)$  ou  $x_i$  est la valeur du caractère  $X$  pour l'individu  $i$ .

## II. Vocabulaires de base



## II. Vocabulaires de base

### 8. L'effectif total

L'effectif total (taille) d'une population ou d'une série statistique (série brute), noté  $N$ , est le nombre d'individus qui composent cette série ou population.

**Exemple** : Au cours d'un semestre, on a observé 20 élèves d'une école dont les notes sont données par le tableau suivant :

1	2	3	4	5	6	7	8	9	10
12	10	12	7	5	12	10	7	13	8
11	12	13	14	15	16	17	18	19	20
7	7	10	12	10	5	5	5	8	13

L'effectif total de cette série est  $N= 20$

## II. Vocabulaires de base

### 9. L'effectif

L'effectif  $n_i$  d'une modalité  $x_i$  est le nombre d'individus ayant cette modalité, ou le nombre de fois que la modalité  $x_i$  est observée.

#### Exemple :

Le classement de la série des 20 notes précédentes par ordre croissant donne les résultats suivants: **5,5,5, 5, 5, 7,7,7, 8,8, 8, 8, 8, 8,10,10,10,10, 10, 10,**

Les modalités obtenues après le regroupement des observations sont: **5,7,8** et **10**.

Donc on a:

La modalité	$x_1 = 5$	a pour effectif	$n_1 = 5$
La modalité	$x_2 = 7$	a pour effectif	$n_2 = 3$
La modalité	$x_3 = 8$	a pour effectif	$n_3 = 6$
La modalité	$x_4 = 10$	a pour effectif	$n_4 = 6$

## II. Vocabulaires de base

**Remarque:**

$$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = 4 + 4 + 2 + 4 + 4 + 2 = \sum_{i=1}^6 n_i = 20$$

**10. L'effectif cumulé croissant**  $N_i$  est le nombre d'individus présentant au plus la modalité  $x_i$ .

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

**Exemple :** Le nombre des élèves qui ont obtenu au plus la note 8 est :

$$N_3 = n_1 + n_2 + n_3 = 4 + 4 + 2 = \sum_{k=1}^3 n_k = 10$$

**11. L'effectif cumulé décroissant**  $N_i^d$  est le nombre d'individus présentant au moins la modalité  $x_k$ .

$$N_i^d = n_i + n_{i+1} + \dots + n_k = \sum_{j=i}^k n_j$$

## II. Vocabulaires de base

**Exemple :** Le nombre des élèves qui ont obtenu au moins la moyenne est :

$$N_4^d = n_4 + n_5 + n_6 = 4 + 4 + 2 = \sum_{k=3}^6 n_k = 10$$

### 12. Fréquence

**La fréquence de la modalité  $x_i$**  est le nombre d'individus possédant ce caractère divisé par l'effectif total de la population. On écrit:

$$f_i = \frac{n_i}{N}$$

#### Remarque

$f_i$  est appelée aussi fréquence relative et on a :

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k$$

#### Exemple :

La fréquence des élèves qui ont obtenu 5 est égale à :  $f_1 = \frac{4}{20} = 0,2$

## II. Vocabulaires de base

### 13. Le Pourcentage ou fréquence en pourcentage d'un caractère

Il s'exprime comme suit:

$$p_i = f_i \times 100$$

**14. Fréquence cumulée croissante** est la proportion d'individus ayant des modalités du caractère étudié inférieures ou égales à  $x_i$ . On a :

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k$$

$$F_3 = f_1 + f_2 + f_3 = 0.2 + 0.2 + 0.1 = \sum_{k=1}^3 f_k = 0.5$$

**Exemple :**

## II. Vocabulaires de base

**15. Fréquence cumulée décroissante** est la proportion d'individus ayant des modalités du caractère étudié supérieures ou égales à  $x_i$ . On a :

$$F_i^d = f_i + f_{i+1} + \dots + f_k = \sum_{j=i}^k f_j$$

**Exemple :**

$$F_3^d = f_3 + f_4 + f_5 + f_6 = 0.1 + 0.2 + 0.2 + 0.1 = \sum_{k=3}^6 f_k = 0.6$$

### 16. Distribution d'un caractère

On appelle distribution d'un caractère X, l'ensemble des couples:

$$\{(x_1, n_1), (x_2, n_2), \dots, (x_k, n_k)\}$$

On peut écrire aussi la distribution sous la forme suivante:  $\{(x_1, f_1), (x_2, f_2), \dots, (x_k, f_k)\}$

# III. Applications Informatiques

# Chapitre 2

## Représentation des données

# I. Tableaux Statistiques

On effectué une enquête sur 20 agriculteurs d'un petit village concernant leurs opinions vis-à-vis un produit agricole appelé GA. Les informations obtenues dans cette enquête sont :

	Nom et prénom	Nombre d'enfants	Revenu (Dh)	Opinion
<b>1</b>	Mohamed Amin	5	3000	Mauvaise (M)
<b>2</b>	Jamal Abid	2	2500	Bien (B)
<b>3</b>	Fouad Rahbi	1	4000	Passable (P)
<b>4</b>	Driss Khilo	4	2500	Très Bien (TB)
<b>5</b>	Mostafa Aydi	1	3000	Assez Bien (AB)
<b>6</b>	Ibrahim Asri	3	2500	Mauvaise (M)
<b>7</b>	Karim Chawki	5	4000	Bien (B)
<b>8</b>	Hamza Figig	4	5300	Assez Bien (AB)

# I. Tableaux Statistiques

	Nom et prénom	Nombre d'enfants	Revenu (Dh)	Opinion
9	Monim Zarwal	6	1500	Très Bien (TB)
10	Ahmed Firgani	3	4000	Passable (P)
11	Abderhami Atar	1	3000	Bien (B)
12	Mohamed Rami	4	2000	Mauvaise (M)
13	Abdhamid mohi	2	2000	Assez Bien (AB)
14	Hamid Chifa	0	5000	Bien (B)
15	Tawfik Micram	5	6500	Très Bien (TB)
16	Bilala Talid	0	5300	Passable (P)
17	Youssef Swalam	4	6500	Mauvaise (M)
18	Ziyad Amo	2	1500	Passable (P)
19	Tarik Filil	1	1200	Bien (B)
20	Bouselham Malok	6	1500	Mauvaise (M)

# I. Tableaux Statistiques

## 1. Distribution d'un caractère

### 1.1 Cas d'un caractère qualitatif

Soit  $X$  un caractère qui désigne l'opinion des agriculteurs par rapport à un produit agricole appelé GA.

Les 5 valeurs possibles de ce caractère sont :  $\{x_1 = M, x_2 = P, x_3 = AB, x_4 = B, x_5 = TB\}$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
M	B	P	TB	AB	M	B	AB	TB	P	B	M	AB	B	TB	P	M	P	B	M

Donc la distribution de ce caractère est:

$$\{(x_1, n_1), (x_2, n_2), \dots, (x_6, n_6)\} = \{(M, 5), (P, 4), (AB, 3), (B, 5), (TB, 3)\}$$

# I. Tableaux Statistiques

Le tableau statistique de ce caractère est:

i	Modalité $x_i$	Effectif $n_i$	Effectif Cumulé Croissant $N_i$	Fréquence $f_i$	Fréquence Cumulé Croissante $F_i$
1	M	5	5	0,25	0,25
2	P	4	9	0,2	0,45
3	AB	3	12	0,15	0,6
4	B	5	17	0,25	0,85
5	TB	3	20	0,15	1
<b>Total N = 20</b>		20		1	

$$f_i = \frac{n_i}{N}$$

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k$$

# I. Tableaux Statistiques

Le tableau statistique de ce caractère est:

i	Modalité $x_i$	Effectif $n_i$	Effectif Cumulé Décroissant $N_i^d$	Fréquence $f_i$	Fréquence Cumulé Décroissante $F_i^d$
1	M	5	20	0,25	1
2	P	4	15	0,2	0,75
3	AB	3	11	0,15	0,55
4	B	5	8	0,25	0,4
5	TB	3	3	0,15	0,15
<b>Total N = 20</b>		20		1	

$$f_i = \frac{n_i}{N}$$

$$N_i^d = n_i + n_{i+1} + \dots + n_k = \sum_{j=i}^k n_j$$

$$F_i^d = f_i + f_{i+1} + \dots + f_k = \sum_{j=1}^k f_j$$

# I. Tableaux Statistiques

Le tableau complet de la distribution de ce caractère est:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$	$N_i^d$	$F_i^d$
M	5	5	0,25	0,25	20	1
P	4	9	0,2	0,45	15	0,75
AB	3	12	0,15	0,6	10	0,65
B	5	17	0,25	0,85	7	0,4
TB	3	20	0,15	1	3	0,15
<b>Total</b>	20		1			

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

$$N_i^d = n_i + n_{i+1} + \dots + n_k = \sum_{j=i}^k n_j$$

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k$$

$$F_i^d = f_i + f_{i+1} + \dots + f_k = \sum_{j=1}^k f_j$$

# I. Tableaux Statistiques

## 1.2 Cas d'un caractère quantitatif

### A- Variable discrète

Soit  $X$  un caractère qui désigne le nombre d'enfants par ménage pour les 20 agriculteurs. Les 7 valeurs possibles de ce caractère sont:

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

Alors le tableau statistique de ce caractère est :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	2	1	4	1	3	5	4	6	3	1	4	2	0	5	0	4	1	2	4

Alors la distribution de ce caractère est:

$$\{(x_1, n_1), (x_2, n_2), \dots, (x_6, n_6)\} = \{(0, 2), (1, 4), (2, 3), (3, 2), (4, 5), (5, 3), (6, 1)\}$$

# I. Tableaux Statistiques

Le tableau statistique de ce caractère est:

i	Modalité $x_i$	Effectif $n_i$	Effectif Cumulé Croissant $N_i$	Fréquence $f_i$	Fréquence Cumulé Croissante $F_i$
1	0	2	2	0,1	0,1
2	1	4	6	0,2	0,3
3	2	3	9	0,15	0,45
4	3	2	11	0,1	0,55
5	4	5	16	0,25	0,8
6	5	3	19	0,15	0,95
7	6	1	20	0,05	1
<b>Total N = 20</b>		20		1	

$$f_i = \frac{n_i}{N}$$

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k$$

# I. Tableaux Statistiques

Le tableau statistique de ce caractère est:

i	Modalité $x_i$	Effectif $n_i$	Effectif Cumulé Décroissant $N_i^d$	Fréquence $f_i$	Fréquence Cumulé Décroissante $F_i^d$
1	0	2	20	0,1	1
2	1	4	18	0,2	0,9
3	2	3	14	0,15	0,7
4	3	2	11	0,1	0,55
5	4	5	9	0,25	0,45
6	5	3	4	0,15	0,2
7	6	1	1	0,05	0,05
<b>Total N = 20</b>		20		1	

$$f_i = \frac{n_i}{N}$$

$$N_i^d = n_i + n_{i+1} + \dots + n_k = \sum_{j=i}^k n_j$$

$$F_i^d = f_i + f_{i+1} + \dots + f_k = \sum_{j=i}^k f_j$$

# I. Tableaux Statistiques

Le tableau complet de la distribution de ce caractère est:

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$	$N_i^d$	$F_i^d$
0	2	2	0,1	0,1	20	1
1	4	6	0,2	0,3	18	0,9
2	3	9	0,15	0,45	14	0,7
3	2	11	0,1	0,55	11	0,55
4	5	16	0,25	0,8	9	0,45
5	3	19	0,15	0,95	4	0,2
6	1	20	0,05	1	1	0,05
<b>Total</b>	20		1			

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{k=1}^i f_k$$

$$N_i^d = n_i + n_{i+1} + \dots + n_k = \sum_{j=i}^k n_j$$

$$F_i^d = f_i + f_{i+1} + \dots + f_k = \sum_{j=i}^k f_j$$

# I. Tableaux Statistiques

## 1.2 Cas d'un caractère quantitatif

### B- Variable continue

Soit  $X$  un caractère continu dont les  $n$  modalités peuvent être regroupées en plusieurs classes  $[e_{i-1}, e_i[$ ,  $1 \leq i \leq n$ .

Soit  $k$  le nombre de classes du caractère continu. Ce nombre est déterminé par la formule Sturges suivante:

$$k = 1 + 3.22 \log(n)$$

Dans la pratique on prend un entier très proche de  $k$ .

# I. Tableaux Statistiques

Ainsi on a:

○  $E = x_{\max} - x_{\min}$  : est l'étendue de la série statistique.

○  $e = \frac{E}{k}$  : est l'étendu de la classe

où  $x_{\max}$  et  $x_{\min}$  étant la valeur maximale et la valeur minimale prises par X.

Alors:

$$e_0 = x_{\min}$$

$$e_1 = x_{\min} + e$$

.....

$$e_k = x_{\min} + ke$$

Donc:

$$\underbrace{[e_0, e_1[}_{1^{\text{ère}} \text{ classe}}$$

$$\underbrace{[e_1, e_2[}_{2^{\text{ème}} \text{ classe}}$$

.....

$$\underbrace{[e_{k-1}, e_k[}_{k^{\text{ème}} \text{ classe}}$$

# I. Tableaux Statistiques

Le tableau statistique relatif à la variable continue peut être représenté comme suit:

N° de classe	Classes	$n_i$	$f_i$
1	$[e_0, e_1[$	$n_1$	$f_1$
2	$[e_1, e_2[$	$n_2$	$f_2$
.	.	.	.
i	$[e_{i-1}, e_i[$	$n_i$	$f_i$
.	.	.	.
k	$[e_{k-1}, e_k[$	$n_k$	$f_k$
Total	-	N	1

# I. Tableaux Statistiques

## Exemple:

Considérons les données suivantes qui désignent les tailles de 20 étudiants:

1,73 1,82 1,87 1,75 1,68 1,64 1,88 1,86 1,89 1,59 **1,5**  
1,67 1,81 1,76 1,72 1,65 1,79 1,89 1,81 **1,9**

L'application de la formule Sturges donne:

$$k = 1 + 3.22 \log(20) \approx 5$$

L'étendu de chaque classe est donné par:

$$e = \frac{E}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{1.9 - 1.5}{5} = 0.08$$

# I. Tableaux Statistiques

Alors:

$$e_0 = 1.5$$

$$e_1 = 1.5 + 0.08 = 1.58$$

$$e_2 = 1.5 + 2 \times 0.08 = 1.66$$

$$e_3 = 1.5 + 3 \times 0.08 = 1.74$$

$$e_4 = 1.5 + 4 \times 0.08 = 1.82$$

$$e_5 = 1.5 + 5 \times 0.08 = 1.90$$

Donc :

$$\underbrace{[1.5, 1.58[}_{1^{\text{ère}} \text{ classe}}$$

$$\underbrace{[1.58, 1.66[}_{2^{\text{ème}} \text{ classe}}$$

$$\underbrace{[1.66, 1.74[}_{3^{\text{ème}} \text{ classe}}$$

$$\underbrace{[1.74, 1.82[}_{4^{\text{ème}} \text{ classe}}$$

$$\underbrace{[1.82, 1.9[}_{5^{\text{ème}} \text{ classe}}$$

# I. Tableaux Statistiques

Soit la série statistique de la variable suivante:

1,73 1,82 1,9 1,75 1,68 1,64 1,88 1,86 1,89 1,59 **1,5** 1,67  
1,81 1,76 1,72 1,65 1,79 1,89 1,81 **1,9**

Le tableau statistique relatif à cette variable est:

N° de classe	Classes	$n_i$	$f_i$
1	[1.5,1.58 [	1	0.05
2	[1.58,1.66[	3	0.15
3	[1.66,1.74[	4	0.2
4	[1.74,1.82[	4	0.2
5	[1.82,1.90]	8	0.4
Total	-	<b>20</b>	<b>1</b>

# I. Tableaux Statistiques

Soit  $X$  une variable qui désigne « **Le Revenu** » des agriculteurs qui est supposée continue. On a:

<b>Revenu X</b>	1200	1500	2000	2500	3000	4000	5000	5300	6500
<b>Effectif <math>n_i</math></b>	1	3	2	3	3	3	1	2	2

Regroupant les revenus en 4 classes de même amplitude comme indiqué par le tableau suivant. Alors on a :

<b>Classes ( <math>10^2</math> )</b>	<b><math>n_i</math></b>	<b><math>N_i</math></b>	<b><math>N_i^d</math></b>	<b><math>f_i</math></b>	<b><math>F_i</math></b>	<b><math>F_i^d</math></b>
[0,20[	4	4	20	0,2	0,2	1
[20,40[	8	12	16	0,4	0,6	0,8
[40,60[	6	18	8	0,3	0,9	0,4
[60,80[	2	20	2	0,1	1	0,1
Total	20			1		

# II. Représentations Graphiques

Le graphique est une traduction visuelle de l'information représenté par un caractère.

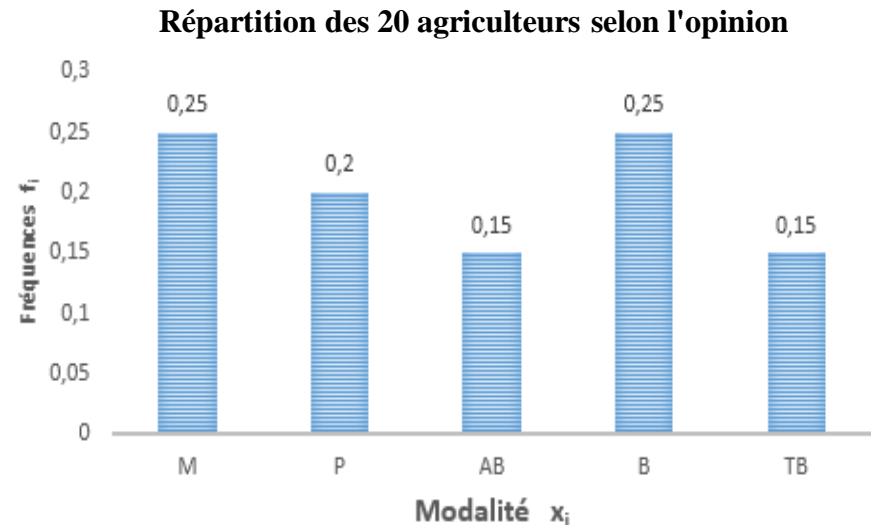
## 1. Cas du caractère qualitatif

### 1.1 Diagramme en barre

Ce diagramme est formé de rectangles d'hauteurs proportionnelles aux effectifs (ou fréquences) des modalités associées.

Ces rectangles ont les mêmes bases et sont constants.

Modalité $x_i$	$n_i$	$f_i$
M	5	0,25
P	4	0,2
AB	3	0,15
B	4	0,25
TB	3	0,15



# II. Représentations Graphiques

## 1.2 Diagrammes circulaires

Ce diagramme est formé de disque dont les secteurs sont proportionnels aux effectifs  $n_i$  (ou fréquences  $f_i$ ) des modalités associées.

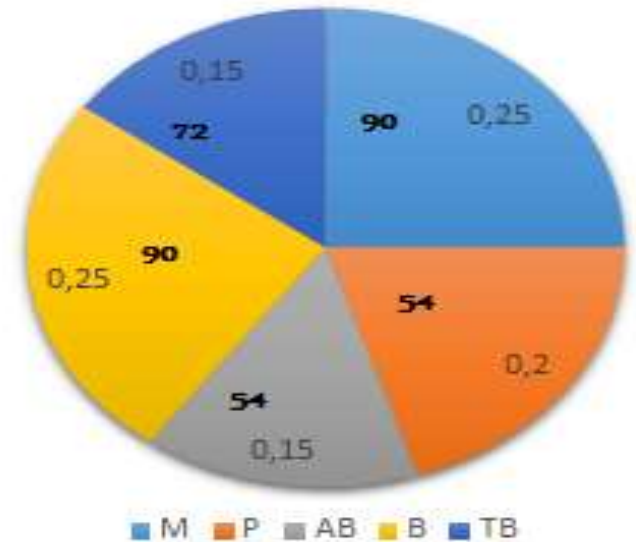
En terme d'angle, on associe à chaque modalité  $x_i$  l'angle  $\theta_i$  du secteur  $i$  tel que:

$$\theta_i = \frac{360^\circ \times n_i}{n} = f_i \times 360^\circ$$

Modalité $x_i$	$n_i$	$f_i$	$\theta_i$
M	5	0,25	90
P	4	0,2	72
AB	3	0,15	54
B	4	0,25	90
TB	3	0,15	54



Répartition des 20 agriculteurs selon l'opinion



# II. Représentations Graphiques

## 2. Cas du caractère quantitatif

### 2.1 Variable discrète

#### A) Diagramme en bâtons

Ce diagramme est formé de segments dont chaque modalité  $x_i$  est associée à un segment d'hauteur  $h_i$  proportionnel à l'effectif  $n_i$  ou à la fréquence  $f_i$ .

Modalité $x_i$	$n_i$	$f_i$
0	2	0,1
1	4	0,2
2	3	0,15
3	2	0,1
4	5	0,25
5	3	0,15
6	1	0,05



# II. Représentations Graphiques

## C. Diagramme cumulatif

Ce diagramme est formé des paliers qui sont horizontaux et les fréquences cumulées croissantes  $F(x)$  qui sont constantes sur chaque intervalle  $[x_{i-1}, x_i[$

tel que:

$$x < x_1 \Rightarrow F(x) = 0$$

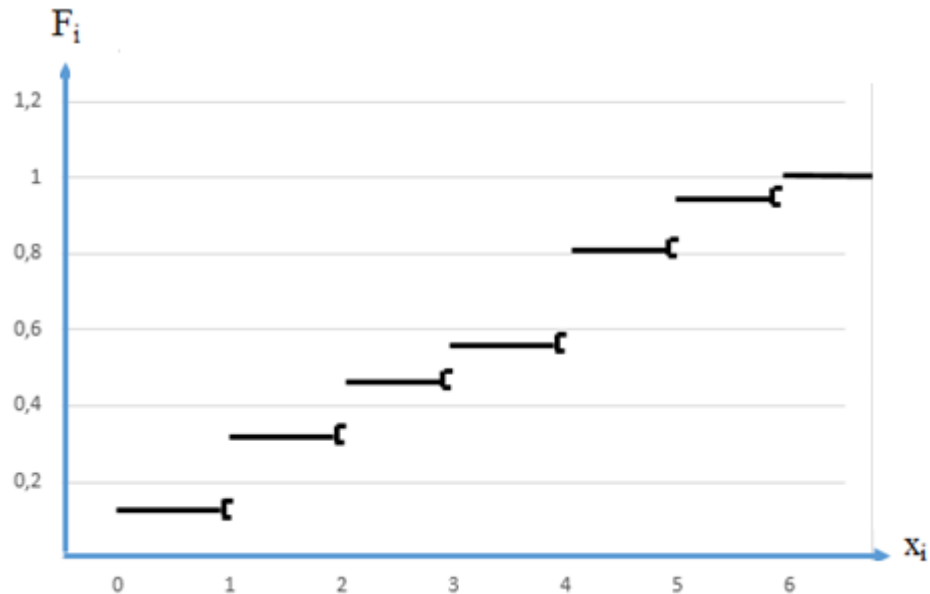
$$x \geq x_n \Rightarrow F(x) = 1$$

et

$$x = x_i \Rightarrow F(x) = F_i = f_1 + \dots + f_i$$

$$x_i \leq x < x_{i+1} \Rightarrow F(x) = F_i$$

Modalité $x_i$	$n_i$	$f_i$	$F_i$
0	2	0,1	0,1
1	4	0,2	0,3
2	3	0,15	0,45
3	2	0,1	0,55
4	5	0,25	0,8
5	3	0,15	0,95
6	1	0,05	1



# II. Représentations Graphiques

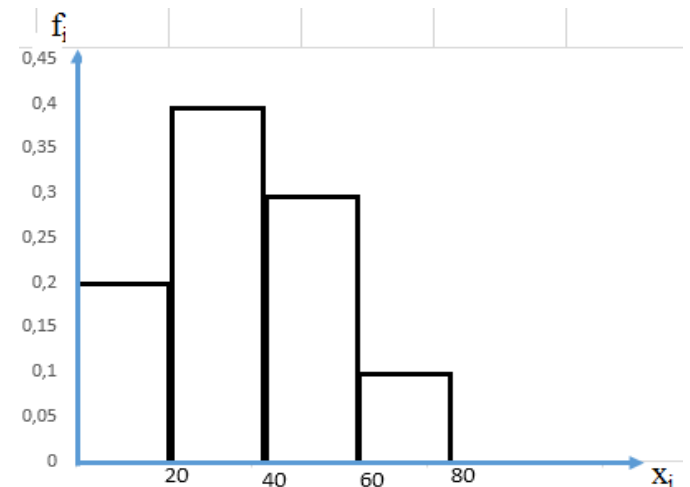
## 2.2 Variable continue

**A) Histogramme :** C'est un diagramme formé de rectangles composés chacun d'une base égale à l'amplitude d'une classe et d'aire proportionnelle à l'effectif (ou fréquence), c'est-à-dire:  $S_i = n_i \times a_i$  ou  $S_i = f_i \times a_i$

### - Cas de classes à amplitude égales

Dans ce cas, l'effectif (fréquence) de la classe  $i$ , noté  $n_i$  (noté  $f_i$ ), est le nombre (proportion) d'individus dont la modalité est comprise entre  $a_i$  et  $a_{i+1}$ .

Classes	Amplitude	$n_i$	$f_i$
[0,20[	20	4	0,2
[20,40[	20	8	0,4
[40,60[	20	6	0,3
[60,80[	20	2	0,1



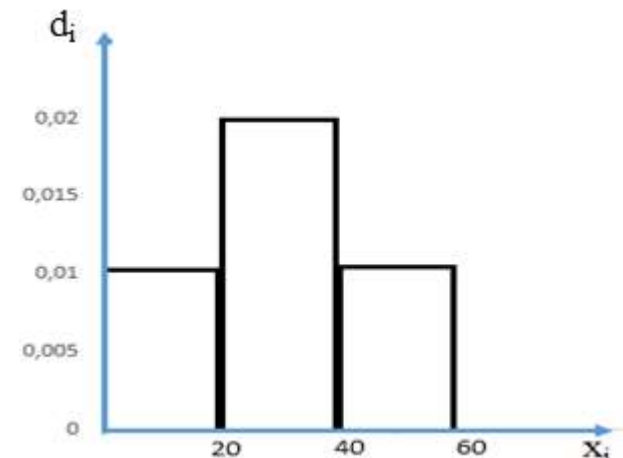
# II. Représentations Graphiques

## - Classes à amplitude inégales

Lorsque les amplitudes sont inégales, il faut effectuer une correction des effectifs (ou fréquences) d'une classe d'effectif  $n_i$  et d'amplitude  $a_i$ .

- L'effectif corrigé, noté  $nc_i$ , est:  $nc_i = \frac{n_i}{a_i}$
- La fréquence corrigée (ou densité de fréquence), notée  $d_i$ , est:  $d_i = \frac{f_i}{a_i}$
- La hauteur de chaque rectangle de l'histogramme est proportionnelle à effectif corrigé (fréquence corrigée). Si en regroupant les deux dernières classes, alors on obtient:

Classes	Amplitude	$f_i$	$d_i$
[0,20[	20	0,2	0,01
[20,40[	20	0,4	0,02
[40,80[	40	0,4	0,01



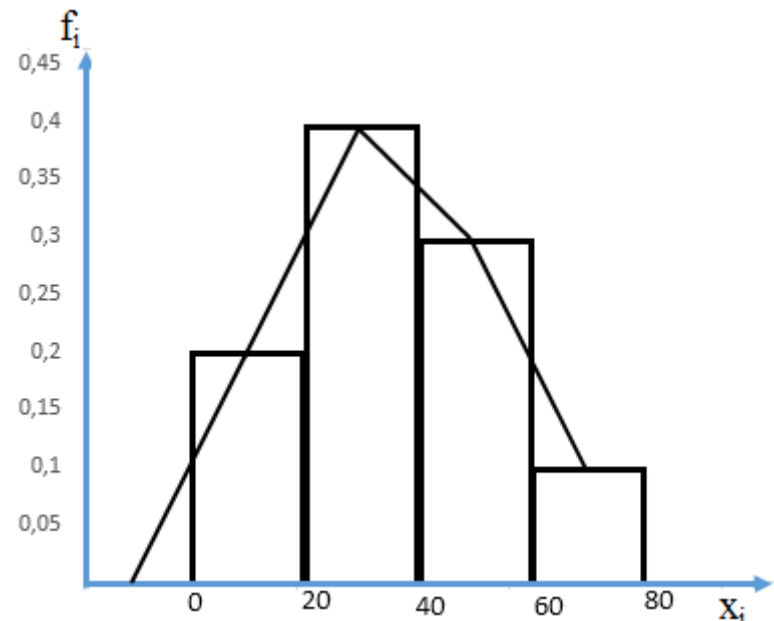
# II. Représentations Graphiques

## B. Polygone de fréquences

Ce diagramme est formé en joignant, par des segments de droites, les milieux des sommets de chaque rectangle de l'histogramme.

Ainsi, il doit commencer par 0 et se terminer par 0.

Classes	$n_i$	$f_i$
[0,20[	4	0,2
[20,40[	8	0,4
[40,60[	6	0,3
[60,80[	2	0,1

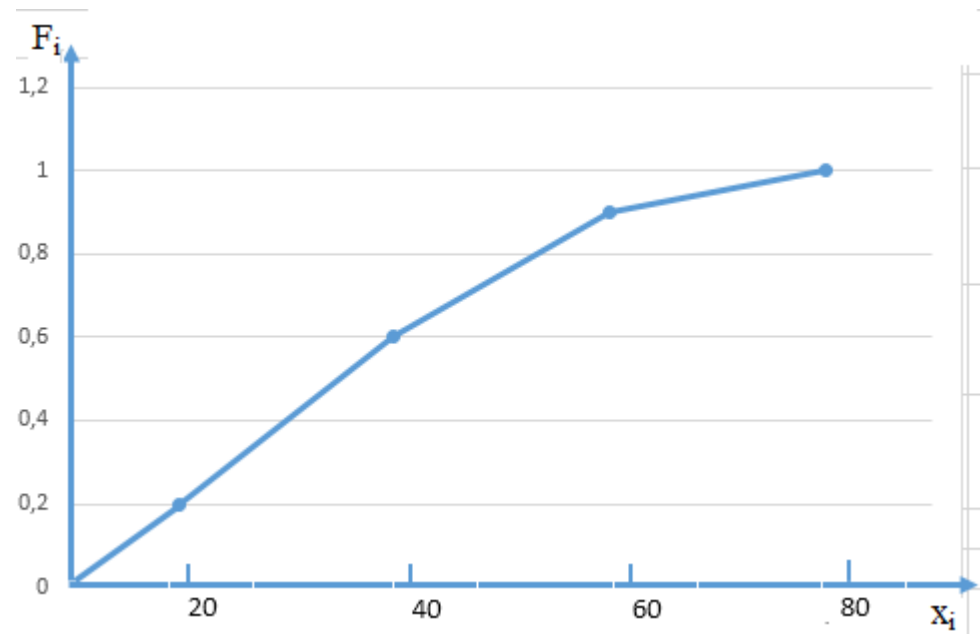


# II. Représentations Graphiques

## C. Diagramme cumulatif

Ce diagramme est formé en joignant les points de coordonnées  $(x_i, F_i)$  pour chaque classe  $[x_{i-1}, x_i[$  où  $F_i$  est la fréquence cumulée de cette classe.

Classes	$n_i$	$f_i$	$F_i$
$[0,20[$	4	0,2	0,2
$[20,40[$	8	0,4	0,6
$[40,60[$	6	0,3	0,9
$[60,80[$	2	0,1	1



# III. Applications Informatiques

# Chapitre 3

## Caractéristiques de Tendance Centrale

# I. Moyennes

## 1. La moyenne arithmétique

### 1.1 La moyenne arithmétique simple

La moyenne arithmétique simple, notée  $\bar{x}$ , d'une série statistique est donnée par :

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

**Exemple :**

Soit la série statistique  $\{10, 25, 20, 25\}$ , alors :

$$\bar{x} = \frac{10 + 25 + 20 + 25}{4} = 20$$

# I. Moyennes

## 1.2 La moyenne arithmétique pondérée

La moyenne arithmétique pondérée d'une variable statistique  $X$  de distribution  $\{(x_i, n_i)_{1 \leq i \leq k}\}$  est donnée par :

□ **Cas d'une variable quantitative discrète :**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i \times x_i = \sum_{i=1}^n f_i \times x_i$$

□ **Cas d'une variable quantitative continue:**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i \times c_i = \sum_{i=1}^n f_i \times x_i$$

où  $c_i = \frac{x_{i+1} + x_i}{2}$  désigne le centre de la classe  $[x_i, x_{i+1}]$

# I. Moyennes

## Exemple : Cas d'une variable quantitative discrète

On reprend le tableau de nombre des enfants des agriculteurs

Modalité $x_i$	$n_i$	$x_i n_i$	$f_i$	$x_i f_i$
0	2	0	0,1	0
1	4	4	0,2	0,8
2	3	6	0,15	0,45
3	2	6	0,1	0,2
4	5	20	0,25	1,25
5	3	15	0,15	0,45
6	1	6	0,05	0,05
Total	20	57	1	2,85

Alors :

$$\bar{x} = \frac{\sum_{i=1}^7 n_i x_i}{n} = \sum_{i=1}^7 f_i x_i = \frac{57}{20} = 2.85$$

# I. Moyennes

## Exemple : Cas d'une variable quantitative continue

On reprend le tableau des revenus des agriculteurs

Classes	$c_i$	$n_i$	$c_i n_i$	$f_i$	$c_i f_i$
[0,20[	10	4	40	0,2	2
[20,40[	30	8	240	0,4	12
[40,60[	50	6	300	0,3	15
[60,80[	70	2	140	0,1	7
Total		20		1	36

$$\bar{x} = \frac{\sum_{i=1}^4 n_i c_i}{n} = \sum_{i=1}^4 f_i c_i = \frac{720}{20} = 36$$

# I. Moyennes

## 2. Moyenne Géométrique

La moyenne géométrique est utilisée dans le calcul du taux d'accroissement moyen et de certains indices statistiques.

La moyenne géométrique, notée  $G$ , d'une variable statistique  $X$  de distribution  $\{(x_i, n_i)_{1 \leq i \leq k}\}$  est donnée par :

$$G = \sqrt[n]{x_1^{n_1} \times \dots \times x_k^{n_k}} \quad \text{où} \quad n = \sum_{i=1}^k n_i$$

Dans la pratique on utilise le logarithme pour calculer  $G$ . En effet:

$$\begin{aligned} \log(G) &= \log\left(\left[x_1^{n_1} \times \dots \times x_k^{n_k}\right]^{\frac{1}{n}}\right) = \frac{1}{n} \log\left(\prod_{i=1}^k x_i^{n_i}\right) \\ &= \frac{1}{n} \sum_{i=1}^k \log(x_i^{n_i}) = \frac{1}{n} \sum_{i=1}^k n_i \log(x_i) = \sum_{i=1}^k f_i \log(x_i) \end{aligned}$$

# I. Moyennes

$$\text{Donc } \log(G) = \frac{1}{n} \sum_{i=1}^k n_i \log(x_i) = \sum_{i=1}^k f_i \log(x_i)$$

**Exemple:** Soit X une variable statistique de distribution

Modalité $x_i$	$n_i$	$n_i \text{Log}(x_i)$
1	2	0
3	1	0,47
4	2	1,2
5	3	1,39
Total	8	0,38

La moyenne géométrique de X est :

$$\log G = \frac{1}{n} \sum_{i=1}^k n_i \log(x_i) = \frac{1}{10} [2 \times \log(1) + 1 \times \log(3) + 2 \times \log(4) + 3 \times \log(5)]$$

$$\log G = 0.38 \Rightarrow G = e^{0.38}$$

# I. Moyennes

## 2. Moyenne harmonique

La moyenne harmonique est utilisée dans le calcul des études du pouvoir d'achat, des durées moyennes, des moyennes de rapports et de pourcentages.

La moyenne harmonique, notée  $H$ , d'une variable statistique  $X$  de distribution  $\{(x_i, n_i)_{1 \leq i \leq k}\}$  est donnée par :

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^k n_i \frac{1}{x_i}} = \frac{1}{\sum_{i=1}^k \frac{f_i}{x_i}} \quad \text{où} \quad n = \sum_{i=1}^k n_i$$

# I. Moyennes

**Exemple:** Soit X une variable statistique de distribution

Modalité $x_i$	$n_i$	$\frac{n_i}{x_i}$
1	2	2
3	1	0,33
4	2	0,5
5	3	0,6
Total	8	3,43

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^k \frac{n_i}{x_i}} = \frac{1}{\sum_{i=1}^k \frac{f_i}{x_i}}$$

La moyenne harmonique de X est :

$$H = \frac{1}{\frac{1}{8} \left( 2 \times \frac{1}{1} + 1 \times \frac{1}{3} + 2 \times \frac{1}{4} + 3 \times \frac{1}{5} \right)} = 2,33$$

$$H = \frac{8}{(2 + 0.3 + 0.5 + 0.6)} = 2,33$$

# I. Moyennes

## 2. Moyenne quadratique

La moyenne quadratique est utilisée dans le calcul de certains paramètres de dispersion.

La moyenne quadratique, notée  $Q$ , d'une variable statistique  $X$  de distribution  $\{(x_i, n_i)_{1 \leq i \leq k}\}$  est donnée par :

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i x_i^2} = \sqrt{\sum_{i=1}^k f_i x_i^2} \quad \text{où} \quad n = \sum_{i=1}^k n_i$$

# I. Moyennes

**Exemple:** Soit X une variable statistique de distribution

Modalité $x_i$	$n_i$	$n_i x_i^2$
1	2	2
3	1	9
4	2	32
5	3	75
Total	8	118

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i x_i^2} = \sqrt{\sum_{i=1}^k f_i x_i^2}$$

La moyenne quadratique de X est :

$$Q = \sqrt{\frac{1}{8} (2 \times 1^2 + 1 \times 3^2 + 2 \times 4^2 + 3 \times 5^2)}$$

$$Q = \sqrt{\frac{1}{8} (2 + 9 + 32 + 75)} = 3.84$$

# I. Moyennes

- La moyenne arithmétique est très sensible aux valeurs extrêmes de la série alors que la moyenne géométrique est peu sensible à ces dernières.
- La moyenne harmonique est plus sensible aux plus petites valeurs de la série qu'aux plus grandes.
- Pour la même série statistique on a :

$$H < G < \bar{x} < Q$$

## II. Mode

Le mode, noté  $M_o$ , d'une série statistique est la valeur qui représente le plus grand effectif ou la fréquence la plus élevée de cette série.

### Remarque:

- Le mode n'est pas nécessairement unique et n'existe pas forcément.
- Pour une variable continue, la classe contenant l'effectif le plus élevé est dite classe modale.

### Détermination du mode

#### 1. Variable statistique qualitative

Considère la variable  $X$  qui désigne “La Couleur préféré” de tableau statistique suivant :

$X_i$	Vert	Rouge	Jaune	Rose
$n_i$	18	14	4	4

## II. Mode

Le mode  $M_0$  est «Vert » car l'effectif associé à ce mode est égal à 18

### 2.Variable statistique quantitative discrète

❑ Soit la série suivante : {5; 7; 6; 7; 3; 1}.

La valeur la plus fréquentée est 7. Donc le mode  $M_0$  est égale à 7.

Dans ce cas on dit qu'il s'agit d'une distribution unimodale

❑ Soit la série suivante : {5; 4; 6; 4; 3; 1;3}.

Les valeurs 3 et 4 sont les valeurs les plus fréquentées .

Dons cette série on a deux modes qui sont 3 et 4.

Dans ce cas on dit qu'il s'agit d'une distribution binomodale

## II. Mode

### 3. Variable statistique quantitative continue

#### - Cas d'amplitudes égales

Soit le tableau des données suivantes:

Classes	$n_i$	Amplitude
[0,20[	4	20
[20,40[	8	20
[40,60[	6	20
[60,80[	2	20

La classe modale est : [20-40[

Pour calculer ce mode il faut appliquer la formule suivante:

$$M_0 = x_1 + \frac{k_1}{k_1 + k_2} (x_2 - x_1)$$

## II. Mode

où :

- ❑  $[x_1, x_2[ = [20, 40[$  : Classe modale (**CM**).
- ❑ L'effectif de la classe modale est égal à 8
- ❑ L'effectif de la classe précédente de la **CM** est égal à 4
- ❑ L'effectif de la classe suivante de la **CM** est égal à 6
- ❑  $k_1 = 8 - 4 = 4$  : la différence entre l'effectif de la classe modale et de la classe précédente de la **CM**.
- ❑  $k_2 = 8 - 6 = 2$  : la différence entre la fréquence (ou l'effectif) de la classe modale et de la classe suivante de la **CM**.

Donc :

$$M_0 = 20 + \frac{4}{4 + 2} (40 - 20) \approx 33,33$$

## II. Mode

### - Cas d'amplitude inégales

Soit le tableau des données suivantes:

Classes	$n_i$	$f_i$	Amplitude	Densité en %
[6,9[	7	0,467	3	15,6
[9,11[	5	0,333	2	16,7
[11,14[	3	0,2	3	6,7

La classe modale  $c$  est la classe ayant l'effectif corrigé ( ou la densité) le(a) plus élevé(e). Donc la classe modale est [9-11[

## II. Mode

où :

- ❑  $[x_1, x_2[ = [9, 11[$  : Classe modale (**CM**).
- ❑ La densité de la classe modale est égal à 16,7
- ❑ La densité de la classe précédente de la **CM** est égal à 15,6
- ❑ La densité de la classe suivante de la **CM** est égal à 6,7
- ❑  $k_1 = 16,7 - 15,6 = 1,1$  : la différence entre la densité de la classe modale et celle de la classe précédente de la **CM**.
- ❑  $k_2 = 16,7 - 6,7 = 10$  : la différence entre la densité de la classe modale et de celle de la classe suivante de la **CM**.

Donc :

$$M_0 = 9 + \frac{1,1}{1,1 + 10} (11 - 9) = 9,2$$

# III. Médiane

La médiane, notée  $M_e$ , d'une série statistique est la valeur qui partage cette série en deux parties d'effectifs égaux.

## Détermination de la médiane:

### - Cas d'une série brute:

Soit une série de  $n$  observations ordonnées par ordre croissant:  $x_1, \dots, x_n$

#### (a) n est un nombre impair

la valeur médiane est l'observation qui occupe le rang  $(n+1)/2$ .

**Exemple 1** : 16 12 1 9 17 19 13 10 4

On ordonne cette série par ordre croissant: 1 4 9 10 **12** 13 16 17 19

La valeur médiane est 12

### III. Médiane

(b) n est un nombre pair

On arrondit le décimal  $n/2$  à l'entier supérieur .

La valeur médiane est l'observation qui occupe le rang  $n/2$ .

**Exemple 1 :** 16 12 1 9 17 19 18 10 3 7 2 11 15 14

On ordonne cette série par ordre croissant:

1 2 3 7 9 10 **12** **11** 14 15 16 17 18 19

La valeur médiane est 11

# III. Médiane

## - Cas d'une distribution :

### (a) Variable statistique discrète

Soit  $X$  une variable statistique de distribution  $\{(x_i, f_i)_{1 \leq i \leq k}\}$ . On utilise les fréquences cumulées croissantes pour déterminer la médiane de  $X$ .

- Dans le cas où  $F_i = 0.5$ , la modalité  $x_i$  correspond à cette fréquence cumulée.
- Dans le cas contraire, la médiane est la modalité  $x_i$  qui correspond à la plus petite fréquence cumulée dépassant strictement 0.5.

# III. Médiane

## Exemple

Modalité $x_i$	$n_i$	$f_i$	$F_i$
0	2	0,1	0,1
1	4	0,2	0,3
2	3	0,15	0,45
3	2	0,1	0,55
4	5	0,25	0,8
5	3	0,15	0,95
6	1	0,05	1

La valeur médiane est égale à 3

# III. Médiane

## (b) Variable statistique continue

Dans un premier temps on détermine la classe  $[ x_{i-1}, x_i [$  contenant la médiane  $M_e$  telle que  $F_{i-1} \leq 0.5 < F_i$ , puis on détermine  $M_e$  en utilisant l'équation suivante:

$$M_e = x_{i-1} + a_i \times \frac{0.5 - F_{i-1}}{f_i}$$

- $[ x_{i-1}, x_i [$  : Classe médiane (CM) (classe qui contient la médiane)
- $a_i$ : Amplitude de la classe médiane
- $f_i$ : Fréquence de la classe médiane
- $F_{i-1}$ : Fréquence cumulée de la classe précédente de la CM

# III. Médiane

Exemple :

Classes	$c_i$	$n_i$	$f_i$	$F_i$	Amplitude
[0,20[	10	4	0,2	0,2	20
[20,40[	30	8	0,4	0,6	20
[40,60[	50	6	0,3	0,9	20
[60,80[	70	2	0,1	1	20

- [ 20, 40 [ : Classe médiane (CM)
- 20: Amplitude de la classe médiane
- 0,4: Fréquence de la classe médiane
- 0,2: Fréquence cumulée de la classe précédente de la CM

$$M_e = x_{i-1} + a_i \times \frac{0.5 - F_{i-1}}{f_i} \quad \Longrightarrow \quad M_e = 20 + 20 \times \frac{0.5 - 0.2}{0.4} = 35$$

# III. Médiane

**Cas de classe de d'amplitude inégale**

# Chapitre 3

## Caractéristiques de dispersion et de concentration

# 1. Les écarts simples

## 1. L'étendue:

L'étendue, notée  $e$ , d'une série statistique est la différence entre la plus grande et la plus petite valeur de celle-ci.

$$e = x_{\max} - x_{\min}$$

**Exemples :**

**Cas 1: Variable statistique discrète** (nombre d'enfants par ménage)

$$e = x_7 - x_1 = 6 - 0 = 6 \text{ enfants}$$

**Cas 2: Variable statistique continue** (Revenu de l'agriculteur)

$$e = x_6 - x_1 = 80 - 0 = 80$$

# 1. Les écarts simples

## 2. Ecart interquantile

- **Quartiles :**

Ils partagent une série statistique en 4 groupes égaux, chacun représentant 25% des observations.

- **Déciles :**

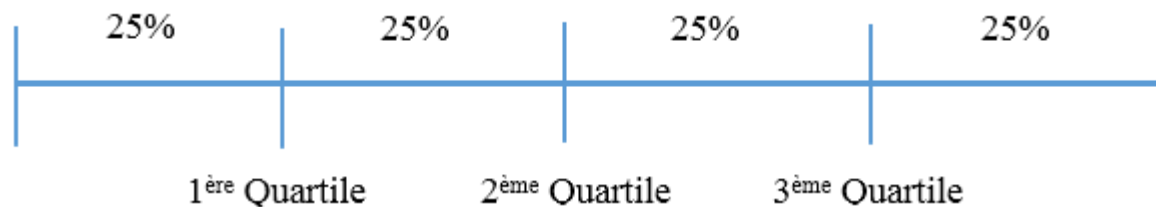
Ils partagent une série statistique en 10 groupes égaux, chacun représentant 10% des observations.

- **Centiles :**

Ils partagent une série statistique en 100 groupes égaux, chacun représentant 1% des observations.

# 1. Les écarts simples

Les quartiles correspondent aux observations pour lesquelles la fréquence cumulée croissante dépasse respectivement 25 %, 50 % et 75%



- **Le premier quartile**, noté  $Q_1$ , d'une série statistique est la plus petite valeur telle qu'au moins 25% des valeurs sont inférieures ou égales à  $Q_1$ .
- **Le deuxième quartile** d'une série statistique c'est la **médiane** ( $Q_2$ )
- **Le troisième quartile**, noté  $Q_3$ , d'une série statistique est la plus petite valeur telle qu'au moins 75% des valeurs sont inférieures ou égales à  $Q_3$ .

# 1. Les écarts simples

## Détermination des quartiles:

### - Cas d'une série brute:

Soit une série statistique de  $n$  observations et  $k$  un nombre tel que  $k = \frac{n}{4}$

### Cas 1: $k$ est un entier

les valeurs des quartiles  $Q_1$  et  $Q_3$ , sont déterminées comme suit:

- Les valeurs sont rangées par ordre croissant.
- $Q_1$  est la valeur de la modalité correspondant au  $1 \times k^{\text{ième}}$  rang
- $Q_3$  est la valeur de la modalité correspondant au  $3 \times k^{\text{ième}}$  rang.

# 1. Les écarts simples

**Exemple 1** : Considérons la série statistique suivante :

17 13 7 27 29 12 19 14 16 1 18 4 8 21 20 11

On range les valeurs de la série par ordre croissant, on obtient:

1 4 7 8 11 12 13 14 16 17 18 19 20 21 27 29

□ Le premier quartile est le  $k^{\text{ème}} = \frac{16}{4} = 4^{\text{ème}}$  rang , c'est-à-dire  $Q_1 = 8$

□ Le troisième quartile est le  $3 \times k^{\text{ème}} = 3 \times \frac{16}{4} = 3 \times 4 = 12^{\text{ème}}$  rang ,  
c'est-à-dire  $Q_3 = 19$

# 1. Les écarts simples

## Cas 1 : $k$ n'est pas un entier

les valeurs des quartiles  $Q_1$  et  $Q_3$ , sont déterminées comme suit:

- a. Les valeurs sont rangées par ordre croissant.
- b. On arrondit le décimal  $1 \times k$  à l'entier supérieur .

$Q_1$  est la valeur de la modalité correspondant au rang de cet entier.

- c. On arrondit le décimal  $3 \times k$  à l'entier supérieur.

$Q_3$  est la valeur de la modalité correspondant au rang de cet entier

# 1. Les écarts simples

**Exemple:** considérons les valeurs rangées par ordre croissant comme suit:

3 4 5 6 7 8 9 10 12 13 15 16 17 19 21 22 23 24 25 27 28 29 39

On a  $n = 23 \Rightarrow k = \frac{n}{4} = 5.75$  et  $3 \times \frac{n}{4} = 17.25$

Donc :

- Le premier quartile est le 6<sup>ème</sup> rang , c'est-à-dire  $Q_1 = 8$
- Le deuxième quartile est le 12<sup>ème</sup> rang , c'est-à-dire  $Q_2 = 16$
- Le troisième quartile est le 18<sup>ème</sup> rang , c'est-à-dire  $Q_3 = 24$

# 1. Les écarts simples

- Cas d'une distribution :

## (b) Variable statistique continue

Dans un premier temps on détermine la classe  $[ x_{i-1}, x_i [$  contenant le quartile  $Q_k$ ,  $k \in \{1, 3\}$ , qui correspond à la plus petite fréquence cumulée dépassant strictement  $\alpha_k \in \{0.25, 0.75\}$  puis on détermine  $Q_k$  en utilisant l'équation suivante:

$$Q_k = x_{i-1} + a_i \times \frac{\alpha_k - F_{i-1}}{f_i} \quad \alpha_1 = 0.25 \text{ et } \alpha_3 = 0.75$$

- $[ x_{i-1}, x_i [$  : Classe qui contient le quartile en question (CQ)
- $a_i$ : Amplitude de la (CQ)
- $f_i$ : Fréquence de la (CQ)
- $F_{i-1}$ :Fréquence cumulé de la classe précédente de la (CQ)

# 1. Les écarts simples

**Exemple** : Revenus des agriculteurs

Classes	$n_i$	$f_i$	$F_i$	Amplitude $a_i$
[0,20[	4	0,2	0,2	20
[20,40[	8	0,4	0,6	20
[40,60[	6	0,3	0,9	20
[60,80[	2	0,1	1	20

$$Q_k = x_{i-1} + a_i \times \frac{\alpha_k - F_{i-1}}{f_i}$$

$$\alpha_1 = 0.25 \text{ et } \alpha_3 = 0.75$$

$Q_1$	$Q_3$
<ul style="list-style-type: none"> <li>- On a <math>\alpha_1 = 0.25</math> (<math>F_0 \leq 0.25 &lt; F_1</math>)</li> <li>- La classe de <math>Q_1</math> est [20,40[</li> <li>- <math>f_i = 0,4</math></li> <li>- <math>F_{i-1} = 0,2</math></li> <li>- <math>Q_1 = 20 + 20 \times \frac{0.25 - 0.2}{0.4} = 22.5</math></li> </ul>	<ul style="list-style-type: none"> <li>- On a <math>\alpha_3 = 0.75</math> (<math>F_2 \leq 0.75 &lt; F_3</math>)</li> <li>- La classe de <math>Q_3</math> est [40,60[</li> <li>- <math>f_i = 0,3</math></li> <li>- <math>F_{i-1} = 0,6</math></li> <li>- <math>Q_3 = 40 + 20 \times \frac{0.75 - 0.6}{0.3} = 50</math></li> </ul>

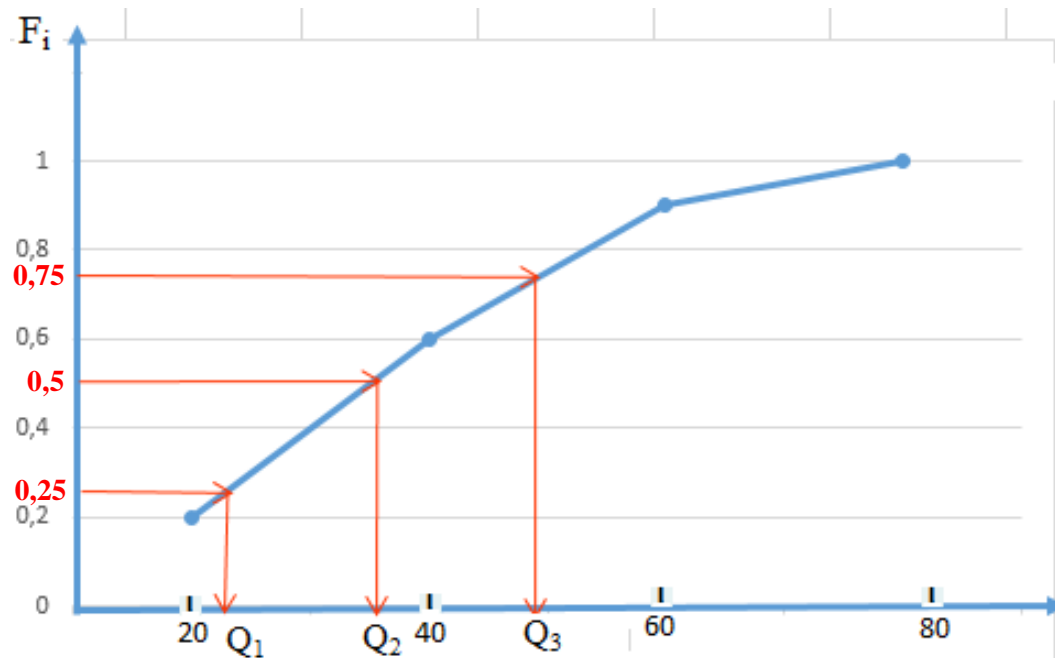
# III. Médiane

**Cas de classe de d'amplitude inégale**

# 1. Les écarts simples

$[Q1; Q3]$  est l'intervalle interquartile, il contient à peu près 50% des valeurs de la série.

On trace le diagramme cumulé et on trouve les quartiles graphiquement comme suit :



## II. Variance et Ecart-type

### 1. Variance

Elle mesure la dispersion des valeurs autour de la moyenne. Plus les valeurs sont regroupées autour de la moyenne, plus la variance est petite. Plus elles sont dispersées, plus la variance est grande.

La variance, notée  $V(X)$ , d'une variable statistique de distribution  $\{(x_i, n_i)_{1 \leq i \leq k}\}$  est donnée par:

$$V(X) = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$$

# II. Variance et Ecart-type

## 2. Ecart-type

L'écart type de  $X$ , noté  $\sigma_X$ , est la racine carrée de la variance.

$$\sigma_X = \sqrt{V(X)}$$

### Remarque:

- Pour une variable statistique continue, les  $x_i$  désignent les centres des classes.
- Dans le cas d'une série brute, la variance de cette série est égale à :

$$V(X) = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 - \bar{x}^2$$

## II. Variance et Ecart-type

### Exemple 1 : Série brute

Considérons la série suivante:  $\{11,14,24,8, 32,9,10,17\}$ .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i = \frac{11+14+24+8+32+9+10+17}{8} = 15,625$$

$$V(X) = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 = \frac{(11-15,625)^2 + (14-15,625)^2 + \dots + (17-15,625)^2}{8} \approx 62,225$$

ou

$$V(X) = \frac{1}{n} \sum_{i=1}^k x_i^2 - \bar{x}^2 = \frac{(11)^2 + (14)^2 + \dots + (17)^2}{8} - (15,625)^2 \approx 62,225$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{62,225} = 7.88$$

## II. Variance et Ecart-type

### Exemple 2: Distribution d'une variable discrète

Nombre d'enfants des agriculteurs

Modalité $x_i$	$n_i$	$n_i x_i$	$x_i^2$	$n_i X_i^2$
0	2		0	0
1	4		1	4
2	3		4	12
3	2		9	18
4	5		16	80
5	3		25	75
6	1		36	36
Total	20		57	225

$$V(X) = \frac{1}{n} \sum_{i=1}^n n_i x_i^2 - \bar{x}^2$$

$$\bar{x} = \frac{(0 \times 2) + (1 \times 4) + (2 \times 3) + \dots + (6 \times 1)}{20}$$

$$= 2.85$$

$$V(x) = \frac{1}{20} (0 + 4 + 12 + \dots + 36) - (2.85)^2$$

$$= 3.1275$$

$$\sigma_x = \sqrt{V(x)} = \sqrt{3.1275} = 1.76$$

## II. Variance et Ecart-type

### Exemple 3: Distribution d'une variable continue

Revenus des agriculteurs répartis en classes

Revenus	$c_i$	$n_i$	$c_i n_i$	$(c_i)^2 n_i$
[0,20[	10	4	40	400
[20,40[	30	8	240	7200
[40,60[	50	6	300	15000
[60,80[	70	2	140	9800
Total		20	720	32400

$$V(X) = \frac{1}{n} \sum_{i=1}^n n_i c^2 - \bar{c}^2$$

$$\bar{c} = \frac{720}{20} = 36$$

$$\begin{aligned} V(x) &= \frac{32400}{20} - (36)^2 \\ &= 324 \end{aligned}$$

$$\sigma_X = \sqrt{V(x)} = \sqrt{324} = 18$$

### III. Coefficient de Variation

Le coefficient de variation, noté CV, d'une variable statistique de moyen  $\bar{x}$  et d'écart-type  $\sigma_X$  est égal au rapport suivant:

$$CV = \frac{\sigma_X}{\bar{x}}$$

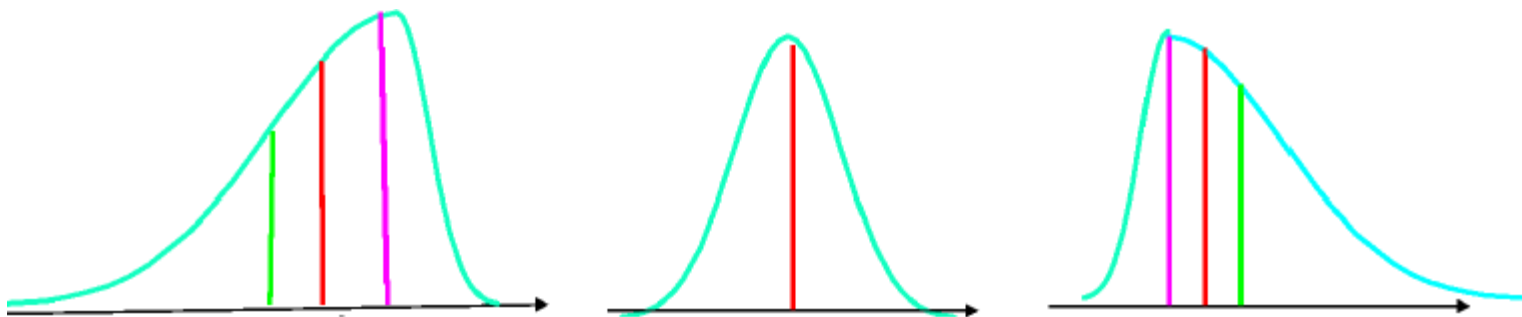
Lorsque deux séries (ou plusieurs) sont exprimées en unités différentes, l'analyse de la dispersion doit se faire par le biais du coefficient de variation

## VI. Forme d'une distribution

La distribution est symétrique si son histogramme est approximativement symétrique par rapport à la droite passant par la médiane. Dans ce cas on a :

$$M_e = M_o = \bar{X}$$

Dans le cas contraire on parle d'une distribution symétrique à gauche si la moitié gauche de son histogramme est plus allongée que sa moitié droite, sinon on parle d'une distribution symétrique à droite



# VI. Forme d'une distribution

## Mesures de l'asymétrie

L'étude de l'asymétrie se fait à travers 3 coefficients qui sont:

➤ **Le coefficient de Yule:** 
$$C_Y = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1}$$

Si  $C_Y = 0$  alors la distribution est symétrique.

➤ **Le coefficient de Pearson:** 
$$C_p = \frac{\bar{X} - M_o}{\sigma}$$

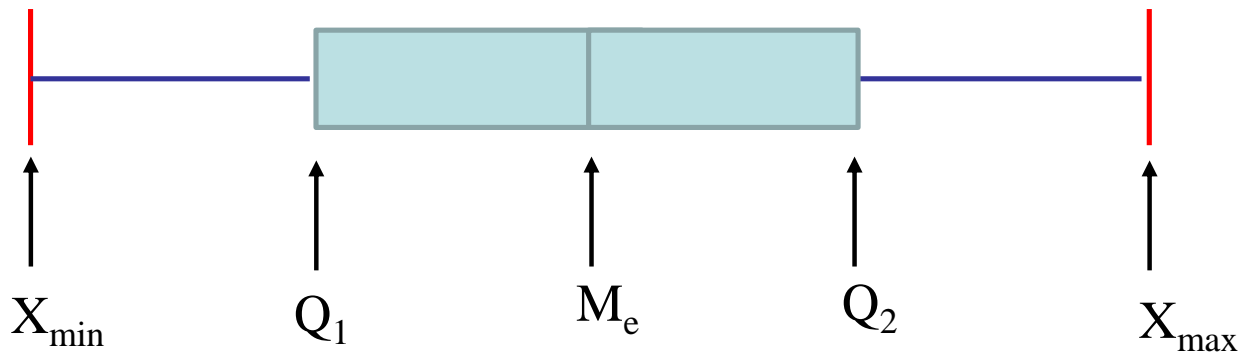
Si  $C_p = 0$  alors la distribution est symétrique.

➤ **Le coefficient de Fisher:** 
$$C_F = \frac{m_3}{\sigma^3} \quad \text{où} \quad m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^3$$

Si  $C_F = 0$  alors la distribution est symétrique.

# V. Boite à Moustaches

C'est un graphique qui permet de représenter les valeurs extrêmes, la médiane, les quartiles  $Q_1$  et  $Q_2$  d'une série statistique, ce qui peut décrire les propriétés d'un échantillon comme la position, la variabilité et l'asymétrie.



# 1. Masse

Expliquer l'utilité de la masse

# 1. Masse

Soit  $X$  une variable continue dont les valeurs sont regroupées en  $k$  classes comme suit:

Classes	Centre $x_i$	Effectif $n_i$
$[e_0, e_1[$	$x_1$	$n_1$
$[e_1, e_2[$	$x_2$	$n_2$
.	.	.
$[e_{k-1}, e_k[$	$x_k$	$n_k$

La masse  $m_i$  et la masse relative  $q_i$  (la proportion de la masse totale) de la variable  $X$  dans la classe  $[e_{i-1}, e_i[$  ou relative à  $x_i$  d'effectif  $n_i$  sont respectivement:

$$\bullet \quad m_i = n_i x_i \qquad \bullet \quad q_i = \frac{m_i}{m}$$

où  $m = \sum_{i=1}^k m_i = \sum_{i=1}^k n_i x_i$  désigne la masse totale de  $X$ .

# 1. Masse

La masse cumulée  $M_i$  et la masse cumulée relative  $Q_i$  associée à la classe  $[e_{i-1}, e_i[$  sont respectivement:

$$\bullet M_i = \sum_{k=1}^i n_k x_k$$

$$\bullet Q_i = \frac{M_i}{\sum_{j=1}^k n_j x_j}$$

**Remarques :**

- i) On a ,  $0 \leq Q_i \leq 1$  pour tout  $i$
- ii) On a  $Q_i = Q(e_i)$

# 1. Masse

## Exemple :

Revenus des agriculteurs répartis en classes

Revenus	$c_i$	$n_i$	$m_i$	$M_i$	$q_i$	$Q_i$
[0,20[	10	4	40	40	0,056	0,056
[20,40[	30	8	240	280	0,333	0,389
[40,60[	50	6	300	580	0,417	0,805
[60,80[	70	2	140	720	0,195	1
Total		$n = 20$	$m = 720$			

- $m_i = n_i x_i$

- $M_i = \sum_{k=1}^i n_k x_k$

- $q_i = \frac{m_i}{m}$

- $Q_i = \frac{M_i}{\sum_{j=1}^k n_j x_j}$

## 2. Médiale

La Médiale, notée  $M_I$ , d'une série statistique est la valeur du caractère ou de la variable  $X$  qui partage la masse totale  $m$  de  $X$  en deux parties égales .

**Remarque:**  $Q(M_I) = 0.5$

Soit  $X$  une variable statistique de distribution  $\{(x_i, f_i)_{1 \leq i \leq k}\}$  . On utilise les masses cumulées relatives pour déterminer la médiale de  $X$ .

Ainsi, la médiale est la modalité  $x_i$  qui correspond à la plus petite masse cumulée relative dépassant strictement 0.5.

## 2. Médiale

### Détermination de la médiale

Dans un premier temps on détermine la classe  $[ x_{i-1}, x_i [$  contenant la médiale  $M_I$  telle que  $Q_{i-1} \leq 0.5 < Q_i$ , puis on détermine  $M_I$  en utilisant l'équation suivante:

$$M_I = x_{i-1} + a_i \times \frac{0.5 - Q_{i-1}}{q_i}$$

- $[ x_{i-1}, x_i [$  : Classe contenant la médiale ( $CM_I$ )
- $a_i$ : Amplitude de la classe  $CM_I$
- $q_i$ : Masse relative associée à la classe  $CM_I$
- $Q_{i-1}$ : Masse cumulée relative associée à la classe précédente de la  $CM_I$

## 2. Médiale

**Exemple :** Revenus des agriculteurs répartis en classes

Revenus	$c_i$	$n_i$	$m_i$	$q_i$	$Q_i$
[0,20[	10	4	40	0,056	0,056
[20,40[	30	8	240	0,333	0,389
<b>[40,60[</b>	<b>50</b>	<b>6</b>	<b>300</b>	<b>0,417</b>	<b>0,805</b>
[60,80[	70	2	140	0,195	1
Total		$n = 20$	$m = 720$		

- [ 40, 60 [ : Classe contenant la médiale ( $CM_I$ )
- 20: Amplitude de la classe  $CM_I$
- 0,417: Masse relative associée à la classe  $CM_I$
- 0,389: Masse cumulée relative associée à la classe précédente de la  $CM_I$

$$M_I = x_{i-1} + a_i \times \frac{0.5 - Q_{i-1}}{q_i} \quad \Rightarrow \quad M_I = 40 + 20 \times \frac{0.5 - 0.389}{0.417} = 45.32$$

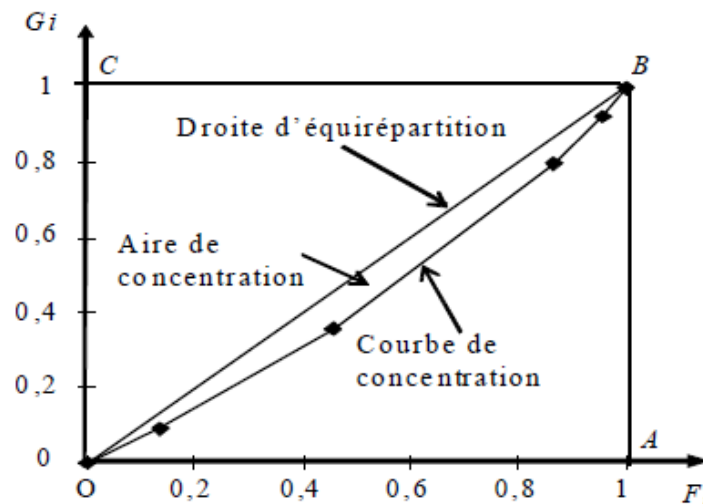
# III. Courbe de concentration

La courbe de concentration ou courbe de Lorenz est obtenue en traçant les points de coordonnées  $(F_i, Q_i)$  en les joignant par des segments de droite dans un repère orthonormé.

On met la fréquence cumulée  $F_i$  sur l'axe des abscisses et la masse cumulée relative  $Q_i$  sur l'axe des ordonnées.

Enfin on trace la droite passant par l'origine d'équation  $Q_i = F_i$ .

**Revoir cette courbe en utilisant les couples  $(F_i, Q_i)$**



# III. Courbe de concentration

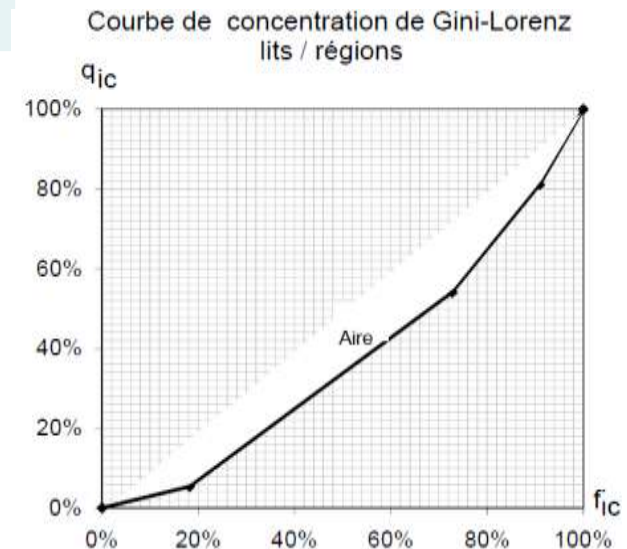
Exemple : Revenus des agriculteurs répartis en classes

Revenus	$c_i$	$n_i$	$f_i$	$F_i$	$m_i$	$q_i$	$Q_i$
[0,20[	10	4	0,2	0,2	40	0,056	0,056
[20,40[	30	8	0,4	0,6	240	0,333	0,389
[40,60[	50	6	0,3	0,9	300	0,417	0,805
[60,80[	70	2	0,1	1	140	0,195	1
Total		$n = 20$	1		$m = 720$		

Revoir cette courbe en utilisant les couples  $(F_i, Q_i)$

Les points de coordonnées  $(F_i, Q_i)$  sont :

$(0,2, 0,056)$  ;  $(0,6, 0,389)$  ;  $(0,9, 0,805)$  ;  $(1, 1)$



# La courbe de concentration

## Formule de calcul de l'indice de Gini

$$G = 1 - \sum_{i=1}^k f_i (Q_{i-1} + Q_i)$$

**Revoir la définition de l'indice de Gini en expliquant son utilité**

- $f_i$  représente la fréquence cumulée à la  $i^{\text{ème}}$  classe.
- $Q_i$  représente la masse cumulée à la  $i^{\text{ème}}$  classe.

## Propriétés de l'indice de Gini

- $0 < G < 1$  ;
- Si  $G$  est proche de 1 alors il y a une forte concentration ;
- Si  $G$  est proche de 0 alors il y a une faible concentration.

# La courbe de concentration

**Exemple :** Revenus des agriculteurs répartis en classes

Revenus	$c_i$	$n_i$	$f_i$	$m_i$	$q_i$	$Q_i$	$S_i = Q_{i-1} + Q_i$	$f_i S_i$
[0,20[	10	4	0,2	40	0,056	0,056	0,056	0,0112
[20,40[	30	8	0,4	240	0,333	0,389	$0,056 + 0,389 = 0,445$	0,178
[40,60[	50	6	0,3	300	0,417	0,805	$0,389 + 0,805 = 1,194$	0,3582
[60,80[	70	2	0,1	140	0,195	1	$0,805 + 1 = 1,805$	0,1805
Total		$n = 20$	1	$m = 720$			3,5	0,7279

$$G = 1 - \sum_{i=1}^k f_i S_i \quad \longrightarrow \quad G = 1 - 0.7279 = 0.2721$$

# Chapitre 4

## Séries doubles

# I. Tableau de contingence

On veut étudier la répartition de 100 étudiants selon le nombre d'année d'obtention de la licence et la mention. Soient :

- X : Nombre d'année d'obtention de la licence
- Y : Mention de la licence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
3	20	10	10	5	45
4	15	10	5	5	35
5	10	5	5	0	20
<b>Marge de Y</b>	45	25	20	10	<b>100</b>

# I. Tableau de contingence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
<b>3</b>	20	10	10	5	<b>45</b>
<b>4</b>	15	10	5	5	<b>35</b>
<b>5</b>	10	5	5	0	<b>20</b>
Marge de Y	45	25	20	10	<b>100</b>



1. La distribution marginale de la variable X est :

X	<b>3</b>	<b>4</b>	<b>5</b>
Marge de X $n_{i\bullet}$	<b>45</b>	<b>35</b>	<b>20</b>

Le nombre  $n_{i\bullet}$  ( $i = 1, 2, 3$ ), est appelé la marge de X ou l'effectif marginal de X.

# I. Tableau de contingence

X\Y	Passable (P)	Assez Bien (AB)	Bien (B)	Très Bien (TB)	Marge de X
3	20	10	10	5	45
4	15	10	5	5	35
5	10	5	5	0	20
Marge de Y	45	25	20	10	100



2. La distribution marginale de la variable Y est :

Y	P	AB	B	TB
Marge de Y $n_{\bullet j}$	45	25	20	10

Le nombre  $n_{\bullet j}$  ( $j = 1, 2, 3, 4$ ), est appelé la marge de Y ou l'effectif marginal de Y.

# I. Tableau de contingence

Soient  $X$  et  $Y$  deux variables statistiques dont les valeurs sont:

$\square x_1, x_2, \dots, x_m$  et  $\square y_1, y_2, \dots, y_k$

Les données de ces deux variables peuvent être regroupées dans un tableau dit tableau de contingence.

$X \backslash Y$	$y_1$	$y_2$	$\dots$	$y_k$	Marge de X $n_{i\bullet}$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1k}$	$n_{1\bullet}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2k}$	$n_{2\bullet}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_m$	$n_{m1}$	$n_{m2}$	$\dots$	$n_{mk}$	$n_{m\bullet}$
Marge de Y $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\dots$	$n_{\bullet k}$	$N$

# I. Tableau de contingence

□ Le nombre  $n_{ik}$  est l'effectif concernant à la fois les modalités  $x_i$  et  $y_j$ , c'est-à-dire la modalité du couple  $(x_i, y_j)$ , où  $i = 1, \dots, m$  et  $j = 1, \dots, k$ .

□  $n_{i\cdot} = \sum_{p=1}^k n_{ip}$  correspond au nombre d'individus ayant les modalités  $x_i$  où  $i = 1, \dots, m$ .

□  $n_{\cdot j} = \sum_{p=1}^m n_{pj}$  correspond au nombre d'individus ayant les modalités  $y_j$  où  $j = 1, \dots, k$ .

□  $N = \sum_{j=1}^k n_{\cdot j} = \sum_{i=1}^m n_{i\cdot} = \sum_{j=1}^k \sum_{k=1}^n n_{kj} = \sum_{i=1}^m \sum_{k=1}^n n_{ik}$  correspond à l'effectif total.

# I. Tableau de contingence

Soient  $X$  et  $Y$  deux variables statistiques qui désignent respectivement le poids (Kg) et la taille (Cm) de 100 animaux.

$X \setminus Y$	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	2	4	2	5	1	14	
[5,10[	4	5	5	2	2	18	
[10,15[	2	10	5	15	4	36	
[15,20[	4	4	2	1	1	12	
[20,25[	5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	

# I. Tableau de contingence

- Fréquence partielle du couple  $(x_i, y_j)$  où  $i = 1, \dots, m$  et  $j = 1, \dots, k$ .  
est donnée par :

$$f_{ij} = \frac{n_{ij}}{n}$$

- Fréquence marginale de  $x_i$  où  $i = 1, \dots, m$  est donnée par:

$$f_{i\bullet} = \frac{n_{i\bullet}}{n}$$

- Fréquence marginale de  $y_j$   $j = 1, \dots, k$  est donnée par:

$$f_{\bullet j} = \frac{n_{\bullet j}}{n}$$

## II. Tableau de fréquences

<b>X\Y</b>	<b>y<sub>1</sub></b>	<b>y<sub>2</sub></b>	<b>...</b>	<b>y<sub>k</sub></b>	<b>Marge de X</b> $f_{i\bullet}$
<b>x<sub>1</sub></b>	$f_{11}$	$f_{12}$	<b>...</b>	$f_{1k}$	$f_{1\bullet}$
<b>x<sub>2</sub></b>	$f_{21}$	$f_{22}$	<b>...</b>	$f_{2k}$	$f_{2\bullet}$
<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	
<b>x<sub>m</sub></b>	$f_{m1}$	$f_{m2}$	<b>...</b>	$f_{mk}$	$f_{m\bullet}$
<b>Marge de Y</b> $f_{\bullet j}$	$f_{\bullet 1}$	$f_{\bullet 2}$	<b>...</b>	$f_{\bullet k}$	<b>1</b>

$$f_{ij} = \frac{n_{ij}}{n}$$

$$f_{i\bullet} = \frac{n_{i\bullet}}{n}$$

$$f_{\bullet j} = \frac{n_{\bullet j}}{n}$$

## II. Tableau de fréquences

**Exemple :**

Le tableau de fréquence du couple  $(x_i, y_j)$  et des fréquences marginales est :

<b>X\Y</b>	<b>P</b>	<b>AB</b>	<b>B</b>	<b>TB</b>	<b>Fréquence Marginale de X</b>
3	0,2	0,1	0,1	0,05	0,45
4	0,15	0,1	0,05	0,05	0,35
5	0,1	0,05	0,05	0	0,2
<b>Fréquence Marginale de Y</b>	0,45	0,25	0,2	0,1	<b>1</b>

## II. Tableau de fréquences

Le tableau de fréquence du couple  $(x_i, y_j)$  et des fréquences marginales est :

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	0,02	0,04	0,02	0,05	0,01	0,14	
[5,10[	0,04	0,05	0,05	0,02	0,02	0,18	
[10,15[	0,02	0,10	0,05	0,15	0,04	0,36	
[15,20[	0,04	0,04	0,02	0,01	0,01	0,12	
[20,25[	0,05	0,02	0,01	0,10	0,2	0,20	
Marge de Y $n_{\bullet j}$	0,17	0,25	0,15	0,33	0,10	N = 1	

### III. Moyennes et Variances Marginales

Soient  $\{x_i : 1 \leq i \leq m\}$  et  $\{y_j : 1 \leq j \leq k\}$  les valeurs discrètes ou les centres de classe pour les variables continues. Alors les moyennes et variances marginales sont:

Variable X	Variable Y
Moyenne marginale de X $\bar{x} = \frac{1}{n} \sum_{i=1}^m n_{i\cdot} x_i = \sum_{i=1}^m f_{i\cdot} x_i$	Moyenne marginale de Y $\bar{y} = \frac{1}{n} \sum_{j=1}^k n_{\cdot j} y_j = \sum_{j=1}^k f_{\cdot j} y_j$
Variance marginale de X $\sigma_X^2 = \frac{1}{n} \sum_{i=1}^m n_{i\cdot} x_i^2 - \bar{x}^2$	Variance marginale de Y $\sigma_Y^2 = \frac{1}{n} \sum_{j=1}^k n_{\cdot j} y_j^2 - \bar{y}^2$

# III. Moyennes et Variances Marginales

Soit l'exemple suivant :

<b>X \ Y</b>	<b>[0,5[</b>	<b>[5,10[</b>	<b>[10,15[</b>	<b>[15,20[</b>	<b>[20,25[</b>	<b>Marge de X</b>	$n_{i\bullet}$
<b>[0,5[</b>	2	4	2	5	1	<b>14</b>	
<b>[5,10[</b>	4	5	5	2	2	<b>18</b>	
<b>[10,15[</b>	2	10	5	15	4	<b>36</b>	
<b>[15,20[</b>	4	4	2	1	1	<b>12</b>	
<b>[20,25[</b>	5	2	1	10	2	<b>20</b>	
<b>Marge de Y</b> $n_{\bullet j}$	17	25	15	33	10		<b>N = 100</b>



Distribution marginale de X

$x_i$	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
<b>2,5</b>	<b>14</b>	35	87,5
<b>7,5</b>	<b>18</b>	135	1012,5
<b>12,5</b>	<b>36</b>	450	5625
<b>17,5</b>	<b>12</b>	210	3675
<b>22,5</b>	<b>20</b>	450	10125
<b>Total</b>	<b>100</b>	<b>1280</b>	<b>20525</b>

# III. Moyennes et Variances Marginales

Calcul de la moyenne marginale et la variance marginale de X

$x_i$	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
2,5	14	35	87,5
7,5	18	135	1012,5
12,5	36	450	5625
17,5	12	210	3675
22,5	20	450	10125
<b>Total</b>	<b>100</b>	<b>1280</b>	<b>20525</b>

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m n_{i\bullet} x_i = \frac{1}{100} \times 1280 = 12.8$$

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^m n_{i\bullet} x_i^2 - \bar{x}^2 = \frac{1}{100} \times 20525 - (12.8)^2 = 205.25 - 163.84 = 41.41$$

## IV. Distributions Conditionnelles

La distribution conditionnelle est la distribution d'une variables sachant que l'autre a une valeur fixe. On écrit  $X/Y = y_j$  (  $X$  sachant  $Y = y_j$  ) ou  $Y/X = x_i$  (  $Y$  sachant  $X = x_i$  )

Il existe deux types de distributions:

- Distribution conditionnelle de  $Y$  sachant  $X = x_i$
- Distribution conditionnelle de  $X$  sachant  $Y = y_j$ .

## IV. Distributions Conditionnelles

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	2	4	2	5	1	14	
[5,10[	4	5	5	2	2	18	
[10,15[	2	10	5	15	4	36	
[15,20[	4	4	2	1	1	12	
[20,25[	5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	

Soient:

- X une variable qui varie en parcourant toutes les classes
- Y est une variable qui fait parti à l'intervalle  $[0, 5[$

La distribution conditionnelle de X sachant que  $Y \in [0, 5[$  et les fréquences conditionnelles obtenues par les effectifs  $n_{i1}$  ( $i = 1, \dots, 5$ ) divisés par  $n_{\bullet 1}$

# IV. Distributions Conditionnelles

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	<b>2</b>	4	2	5	1	14	
[5,10[	<b>4</b>	5	5	2	2	18	
[10,15[	<b>2</b>	10	5	15	4	36	
[15,20[	<b>4</b>	4	2	1	1	12	
[20,25[	<b>5</b>	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	<b>17</b>	25	15	33	10	N = 100	



X\Y	L'effectif conditionnel de $X / Y \in [0,5[$	Fréquence conditionnelle de $X / Y \in [0,5[$
[0,5[	<b>2</b>	0,12
[5,10[	<b>4</b>	0,23
[10,15[	<b>2</b>	0,12
[15,20[	<b>4</b>	0,23
[20,25[	<b>5</b>	0,3
$n_{\bullet 1}$	<b>17</b>	1

# IV. Distributions Conditionnelles

La distribution conditionnelle de Y sachant que  $X \in [10,15[$  et les fréquences conditionnelles obtenues par les effectifs  $n_{3j}$  ( $j = 1, \dots, 5$ ) divisés par  $n_{3\bullet}$ .

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	2	4	2	5	1	14	
[5,10[	4	5	5	2	2	18	
<b>[10,15[</b>	<b>2</b>	<b>10</b>	<b>5</b>	<b>15</b>	<b>4</b>	<b>36</b>	
[15,20[	4	4	2	1	1	12	
[20,25[	5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	N = 100	



X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{3\bullet}$
L'effectif conditionnel de $Y/X \in [10,15[$	<b>2</b>	<b>10</b>	<b>5</b>	<b>15</b>	<b>4</b>	<b>36</b>	
Fréquence conditionnelle de $Y/X \in [10,15[$	0.055	0.277	0.138	0.416	0.111	1	

# V. Moyennes et Variances Conditionnelles

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	Marge de X	$n_{i\bullet}$
[0,5[	2	4	2	5	1	14	
[5,10[	4	5	5	2	2	18	
[10,15[	2	10	5	15	4	36	
[15,20[	4	4	2	1	1	12	
[20,25[	5	2	1	10	2	20	
Marge de Y $n_{\bullet j}$	17	25	15	33	10	<b>N = 100</b>	

La distribution conditionnelle de X sachant [10,15[ est une série univariée obtenue par l'élimination de toutes les classes sauf celle correspond à  $n_{3\bullet}$ .



X\Y	$n_{3\bullet}$
[0,5[	2
[5,10[	5
[10,15[	5
[15,20[	2
[20,25[	1

# V. Moyennes et Variances Conditionnelles

Calcul de la moyenne conditionnelle et la variance conditionnelle de X

X\Y	$c_i$	$n_{3\bullet}$	$c_i n_{3\bullet}$	$c_i^2 n_{3\bullet}$
[0,5[	2,5	2	5	12,5
[5,10[	7,5	5	37,5	281,25
[10,15[	12,5	5	62,5	781,25
[15,20[	17,5	2	35	612,5
[20,25[	22,5	1	22,5	506,25
<b>Total</b>		15	162,5	2193,75

$$\bar{x}_{[10,15]} = \bar{x}_{/3} = \frac{1}{n_{\bullet 3}} \sum_{i=1}^m c_i n_{i3} = \frac{162.5}{15} = 10.83$$

$$\sigma_{/[10,15]}^2 = \sigma_{/3}^2 = \frac{1}{n_{\bullet 3}} \sum_{i=1}^m n_{ij} x_i^2 - (\bar{x}_{/3})^2 = \frac{2193.75}{15} - (10.83)^2 = 34.7$$

# V. Moyennes et Variances Conditionnelles

**Variable X sachant  $Y = y_j$**

Moyenne conditionnelle  
de X sachant  $Y = y_j$

$$\bar{x}_{/j} = \frac{1}{n_{\bullet j}} \sum_{i=1}^m n_{ij} x_i = \sum_{i=1}^m f_{i/j} x_i$$

Variance conditionnelle  
de X sachant  $Y = y_j$

$$\sigma_{X/j}^2 = \frac{1}{n_{\bullet j}} \sum_{i=1}^m n_{ij} (x_i - \bar{x}_{/j})^2$$

**Variable Y sachant  $X = x_i$**

Moyenne conditionnelle  
de Y sachant  $X = x_i$

$$\bar{y}_{/i} = \frac{1}{n_{i\bullet}} \sum_{j=1}^k n_{ij} y_j = \sum_{j=1}^k f_{j/i} y_j$$

Variance conditionnelle  
de Y sachant  $X = x_i$

$$\sigma_{Y/i}^2 = \frac{1}{n_{i\bullet}} \sum_{j=1}^k n_{ij} (y_j - \bar{y}_{/i})^2$$

# VI. Covariance

La covariance de X et Y est la quantité

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\ &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} x_i y_j - \bar{x} \times \bar{y} \\ &= \overline{xy} - \bar{x} \times \bar{y} \end{aligned}$$

## Remarque:

Dans le cas des séries quantitatives non-groupées, on a  $n_{ij} = 1$

# VI. Covariance

Soient X et Y deux variables statistiques qui désignent respectivement le poids (Kg) et la taille (Cm) de 6 animaux.

## Cas 1: Variables quantitatives non-groupées

	Poids ( $x_i$ )	Taille ( $y_i$ )	$x_i y_i$
1	12	10	120
2	14	11	154
3	15	13	195
4	6	7	42
5	9	11	99
6	7	5	35
Total	63	57	645

On a :

$$\bar{x} = \frac{63}{6} = 10.5 \quad \text{et} \quad \bar{y} = \frac{57}{6} = 9,5$$

$$Cov(X, Y) = \frac{645}{6} - (10.5 \times 9,5) = 7.75$$

# VI. Covariance

## Cas 2: Variables quantitatives groupées

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	$x_i$	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
[0,5[	2	4	2	5	1	<b>2,5</b>	<b>14</b>	35	87,5
[5,10[	4	5	5	2	2	<b>7,5</b>	<b>18</b>	135	1012,5
[10,15[	2	10	5	15	4	<b>12,5</b>	<b>36</b>	450	5625
[15,20[	4	4	2	1	1	<b>17,5</b>	<b>12</b>	210	3675
[20,25[	5	2	1	10	2	<b>22,5</b>	<b>20</b>	450	10125
$y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>			<b>1280</b>	<b>20525</b>
$n_{\bullet j}$	<b>17</b>	<b>25</b>	<b>15</b>	<b>33</b>	<b>10</b>		<b>N = 100</b>		
$n_{\bullet j}y_j$	42,5	187,5	187,5	577,5	225	<b>1220</b>			
$n_{\bullet j}y_j^2$	106,25	1406,25	486,75	10106,25	5062,5	<b>17168</b>			

$$\bar{x} = \frac{1280}{100} = 12.8$$

et

$$\bar{y} = \frac{1220}{100} = 12.2$$

$$V(Y) = \frac{17168}{100} - (12.2)^2 = 7.84$$

et

$$V(X) = \frac{20525}{100} - (12.8)^2 = 56.41$$

# VI. Covariance

Calculons la somme suivante:  $\sum_{i=1}^m \left( \sum_{j=1}^k n_{ij} y_j \right) \times x_i$  . Pour chaque case on fixe i et on fait varier j.

Case N°1:

						<b>Centre de X</b> $x_1$
<b>Effectif marginaux</b> $n_{1j}$ de [0,5[	2	4	2	5	1	<b>2,5</b>
<b>Centre de Y</b> $y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (2 \times 2.5 + 4 \times 7.5 + 2 \times 12.5 + 5 \times 17.5 + 1 \times 22.5) \times 2.5 = 425$$

Case N°2:

						<b>Centre de X</b> $x_2$
<b>Effectif marginaux</b> $n_{2j}$ de [5,10[	4	5	5	2	2	<b>7,5</b>
<b>Centre de Y</b> $y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>	

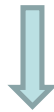


$$(n_{21}y_1 + n_{22}y_2 + n_{23}y_3 + n_{24}y_4 + n_{25}y_5) \times x_2 = (4 \times 2.5 + 5 \times 7.5 + 5 \times 12.5 + 2 \times 17.5 + 2 \times 22.5) \times 7.5 = 1425$$

# VI. Covariance

Case N°3:

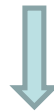
						<b>Centre de X</b> $x_3$
<b>Effectif marginaux</b> $n_{3j}$ de [10,15[	2	10	5	15	4	<b>12,5</b>
<b>Centre de Y</b> $y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (2 \times 2.5 + 10 \times 7.5 + 5 \times 12.5 + 15 \times 17.5 + 4 \times 22.5) \times 12.5 = 6187.5$$

Case N°4:

						<b>Centre de X</b> $x_4$
<b>Effectif marginaux</b> $n_{4j}$ de [15,20[	4	4	2	1	1	<b>17,5</b>
<b>Centre de Y</b> $y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>	



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (4 \times 2.5 + 4 \times 7.5 + 2 \times 12.5 + 1 \times 17.5 + 1 \times 22.5) \times 17.5 = 1837.5$$

# VI. Covariance

Case N°5:

								<b>Centre de X</b> $x_5$
<b>Effectif marginaux</b> $n_{5j}$	<b>de [15,20[</b>	5	2	1	10	2		<b>22,5</b>
<b>Centre de Y</b>	$y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>		



$$(n_{11}y_1 + n_{12}y_2 + n_{13}y_3 + n_{14}y_4 + n_{15}y_5) \times x_1 = (5 \times 2.5 + 2 \times 7.5 + 1 \times 12.5 + 10 \times 17.5 + 2 \times 22.5) \times 22.5 = 5850$$

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^k n_{ij} x_i y_j - \bar{x} \times \bar{y} \\ &= \frac{1}{100} (425 + 1425 + 6187.5 + 1837.5 + 5850) - 12.2 \times 12.8 \\ &= 157.25 - 156.16 = 1.09 \end{aligned}$$

# V. Corrélation

Le coefficient de corrélation linéaire de X et Y est donnée par:

$$R = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$R^2$  désigne le coefficient de détermination de X et Y

## Remarques importante :

- ❑ On a toujours  $-1 \leq R \leq 1 \Rightarrow 0 \leq R^2 \leq 1$ .
- ❑ La covariance dépend des unités ce qui est difficile à interpréter .
- ❑ La corrélation ne dépend pas des unités.

## V. Corrélation

$$\bar{x} = \frac{1280}{100} = 12.8$$

et

$$\bar{y} = \frac{1220}{100} = 12.2$$

$$V(X) = \frac{17168}{100} - (12.8)^2 = 7.84$$

et

$$V(Y) = \frac{20525}{100} - (12.2)^2 = 56.41$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{7.84} = 2.8$$

et

$$\sigma_Y = \sqrt{V(Y)} = \sqrt{56.41} = 7.51$$

$$Cov(X, Y) = 1.09$$

$$R = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1.09}{2.8 \times 7.51} = 0.051$$

# VI. Régression linéaire

Soient X et Y deux variables statistiques dont les valeurs sont:

□  $x_1, x_2, \dots, x_m$  et □  $y_1, y_2, \dots, y_k$

Considérons la série statistique constituée de n couples des valeurs prises par les deux variables pour chaque individu :  $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)$

L'objectif de la régression linéaire est d'étudier la relation qui peut exister entre X et Y en cherchant la fonction  $f$  telle que:  $Y = f(X)$

## VI. Régression linéaire

L'ensemble des poids  $(x_i, y_i)$  qui constituent « un nuage de poids » sont représentés dans un repère orthogonal, ce qui nous permet d'avoir une idée claire sur la nature de la relation entre X et Y et le type de courbe qui ajustera le mieux.

L'ajustement linéaire se base sur le choix de la droite qui passe le plus proche de tous les points du nuage. Cette droite est appelée droite de régression linéaire de Y en X, noté  $D_{Y/X}$ , et son équation est donnée par:  $y = a + bx$   
où  $a$  et  $b$  représentent les paramètres de cette droite.

# VI. Régression linéaire

La détermination de l'équation de la droite de régression linéaire se base sur l'utilisation de la méthode des moindres carrés (M.M.C) qui consiste à minimiser la somme des carrés des distances entre les points du nuage et cette droite.

C'est-à-dire:

$$\text{Minimiser} \left\{ (E) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \right\}$$

Les paramètres  $a$  et  $b$  qui minimisent l'équation (E) sont donnés par :

$$b = \frac{\text{Cov}(X, Y)}{V(Y)} \quad \text{et} \quad a = \bar{y} - b\bar{x}$$

# VI. Régression linéaire

## Coefficient de corrélation linéaire

Ce coefficient mesure le degré de liaison entre  $X$  et  $Y$ . Il est donné par:

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

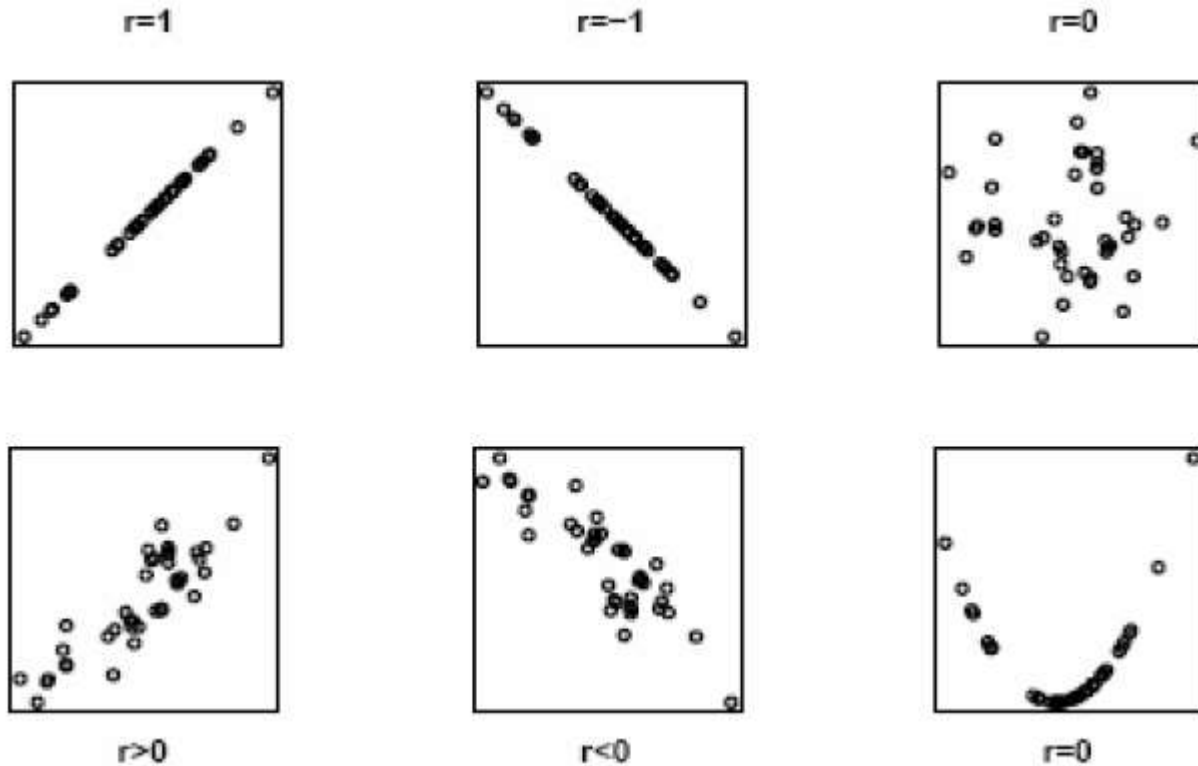
$r^2(X, Y)$  désigne le coefficient de détermination

On a:  $-1 \leq r \leq +1$

- Si  $r < 0$  alors la liaison est négative.
- Si  $r > 0$  alors la liaison est positive.
- Si  $r = 0$  alors la liaison est nulle.

# VI. Régression linéaire

- Plus la valeur absolue de  $r$  est proche de  $1$  plus la liaison entre X et Y est forte.
- Plus la valeur absolue de  $r$  est proche de  $0$  plus la liaison entre X et Y est faible



# VI. Régression linéaire

## Exemple 1:

	Poids ( $x_i$ )	$x_i^2$	Taille ( $y_i$ )	$y_i^2$	$x_i y_i$
1	12	144	10	100	120
2	14	196	11	121	154
3	15	225	13	169	195
4	6	36	7	49	42
5	9	81	11	121	99
6	7	49	5	25	35
Total	63	731	57	585	645

$$b = \frac{Cov(X, Y)}{V(Y)} = \frac{9,25}{7.25} = 1.27$$

$$a = \bar{y} - b\bar{x} = 9.5 - 1.27 \times 10.5 = -3.835$$

Donc  $y = -3.835 + 1.27x$

On a :  $\bar{x} = \frac{63}{6} = 10.5$  ,  $\bar{y} = \frac{57}{6} = 9,5$  et  $Cov(X, Y) = \frac{645}{6} - (10.5 \times 9,5) = 7.75$

$$V(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2 = \frac{731}{6} - (10.5)^2 = 11.58 \quad \text{et} \quad V(Y) = \frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2 = \frac{585}{6} - (9.5)^2 = 7.25$$

# VI. Régression linéaire

## Exemple 2:

X\Y	[0,5[	[5,10[	[10,15[	[15,20[	[20,25[	$x_i$	$n_{i\bullet}$	$n_{i\bullet}x_i$	$n_{i\bullet}x_i^2$
[0,5[	2	4	2	5	1	<b>2,5</b>	<b>14</b>	35	87,5
[5,10[	4	5	5	2	2	<b>7,5</b>	<b>18</b>	135	1012,5
[10,15[	2	10	5	15	4	<b>12,5</b>	<b>36</b>	450	5625
[15,20[	4	4	2	1	1	<b>17,5</b>	<b>12</b>	210	3675
[20,25[	5	2	1	10	2	<b>22,5</b>	<b>20</b>	450	10125
$y_j$	<b>2,5</b>	<b>7,5</b>	<b>12,5</b>	<b>17,5</b>	<b>22,5</b>			<b>1280</b>	<b>20525</b>
$n_{\bullet j}$	<b>17</b>	<b>25</b>	<b>15</b>	<b>33</b>	<b>10</b>		<b>N = 100</b>		
$n_{\bullet j}y_j$	42,5	187,5	187,5	577,5	225	<b>1220</b>			
$n_{\bullet j}y_j^2$	106,25	1406,25	486,75	10106,25	5062,5	<b>17168</b>			

$$\bar{x} = \frac{1280}{100} = 12.8 \quad , \quad \bar{y} = \frac{1220}{100} = 12.2 \quad , \quad V(Y) = \frac{17168}{100} - (12.2)^2 = 7.84 \quad \text{et} \quad Cov(X, Y) = 1.09$$

$$b = \frac{Cov(X, Y)}{V(Y)} = \frac{1,09}{7.84} = 0.14 \quad , \quad a = \bar{y} - b\bar{x} = 12.2 - 0.14 \times 12.8 = 1.79 \quad \text{Donc} \quad y = 1.79 + 0.14x$$