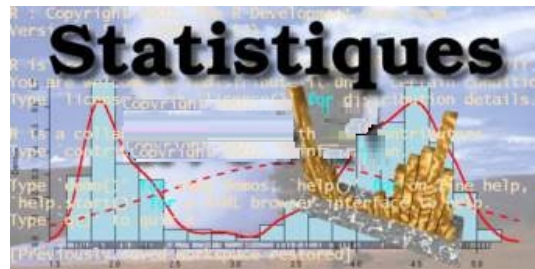


# STATISTIQUE DESCRIPTIVE

*Cours-Mathématique*



Moufida Tabet

# Table of contents



<b>I - Chapitre 1 : Les concepts de base</b>	<b>3</b>
1. Terminologie .....	3
1.1. Statistique .....	3
1.2. Population .....	3
1.3. Échantillon .....	4
1.4. Individu ou unité statistique .....	4
1.5. La taille .....	4
1.6. Le caractère .....	4
1.7. Les modalités .....	4
1.8. Variable qualitatif .....	5
1.9. Variable quantitatif .....	5
1.10. Variable quantitatif discret .....	5
1.11. Variable quantitatif continue .....	5
1.12. Série statistique .....	6
2. Tableau statistique .....	6
2.1. Tableau brut .....	6
2.2. Tableau de contingence. ....	7
2.3. Tableau des fréquences. ....	9

# Chapitre 1 : Les concepts de base

Terminologie	3
Tableau statistique	6

Avant tout calcul statistique, il est nécessaire de disposer de données. Pour atteindre cet objectif, il est impératif de définir très précisément la population sur laquelle s'effectue l'enquête et les variables que l'on cherche à appréhender. Le type de ces variables conditionne les traitements statistiques qu'elles seront susceptibles de subir.

## 1. Terminologie

Statistique	3
Population	3
Échantillon	4
Individu ou unité statistique	4
La taille	4
Le caractère	4
Les modalités	4
Variable qualitatif	5
Variable quantitatif	5
Variable quantitatif discret	5
Variable quantitatif continue	5
Série statistique	6

### 1.1. Statistique

La statistique est une méthode scientifique dont l'objet est de recueillir, d'organiser, de résumer et d'analyser les données d'une enquête, d'une étude ou d'une expérience, aussi bien que de tirer les conclusions logiques et de prendre les décisions qui s'imposent à partir des analyses effectuées.

### 1.2. Population

La population est un ensemble d'individus définis par une propriété commune donnée.

 Example

---

Si l'on veut étudier la durée de vie des ampoules électriques fabriquées par une compagnie, la population considérée est l'ensemble de toutes les ampoules fabriquées par cette compagnie, les étudiants d'une classe, les contribuables français, les ménages lillois. . .

### 1.3. Échantillon

L'échantillon est une sous-ensemble de la population.

 Example

---

Pour établir la durée de vie des ampoules électriques produites par une machine, on peut prélever au hasard un certain nombre d'ampoules - un échantillon- parmi toutes les celles produites par cette machine.

### 1.4. Individu ou unité statistique

Chaque élément de la population ou de l'échantillon appelé individu ou unité statistique.

 Example

---

Dans l'exemple précédant, chaque ampoule constitue un individu ou une unité statistique.

### 1.5. La taille

La taille représente le nombre d'individus d'un échantillon ou d'une population. Elle est symbolisée par « n » dans le cas d'un échantillon et par « N » dans le cas d'une population.

### 1.6. Le caractère

Le caractère c'est l'aspect particulier que l'on désire étudier.


 Example

---

Concernant un groupe de personnes, on peut s'intéresser à leur age, leur sexe leur taille...

### 1.7. Les modalités

Les modalités sont les différentes manières d'être que peut présenter un caractère.

 Example

---

Le sexe est un caractère qui présente deux modalités : féminin ou masculin

 Example

---

Quant au nombre d'enfants par famille, les modalités de ce caractère peuvent être 0,1 2,3...,20.

## 1.8. Variable qualitatif

La variable est dite qualitative quand les modalités sont des catégories c'est a dire ses modalités ne s'expriment pas par un nombre.

- Variable qualitative nominale: La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées.
- Variable qualitative ordinale : La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Le fait de pouvoir ou non ordonner les modalités est parfois discutable. Par exemple : dans les catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession','enseignant', 'artisan', l'ordre devient beaucoup plus discutable.

 Example

---

La religion, le sexe, l'opinion...

## 1.9. Variable quantitatif

Une variable est dite quantitative si toute ses valeurs possibles sont numériques.

 Example

---

L'age, la taille, le poids...

## 1.10. Variable quantitatif discret

L'ensemble des valeurs que peut prendre le variable est fini ou dénombrable. Le plus souvent, ces valeurs sont entières.

 Example

---

Le nombre d'enfant dans une famille, le nombre de téléviseurs par foyer et la pointure des souliers.

## 1.11. Variable quantitatif continue

 Definition

---

Le variable peut prendre théoriquement n'importe quelle valeur dans un intervalle donné de nombres réels.

 Example

---

La taille d'un individu, le poids...

## 1.12. Série statistique

### Definition

Série statistique c'est l'ensemble des différentes données associées à un certain nombre d'individus.

### Example

La série suivante résulte d'une courte enquête auprès de quelques personnes pour connaître leur âge :

18 21 19 19 17 22 27 18 18 17 20 20 23

## 2. Tableau statistique

Tableau brut	6
Tableau de contingence.	7
Tableau des fréquences.	9

Ils constituent le moyen le plus sûr de pouvoir répondre aux questions posées de par leur systématisme. Sauf cas exceptionnels, les données statistiques sont présentées sous forme de tableau. D'une part, cela permet d'appréhender l'information qui est synthétisée et d'autre part facilite ou rend possible les calculs.

### 2.1. Tableau brut

On considère une étude statistique portant sur un échantillon de taille  $N$  individus. On mesure plusieurs variables statistiques. La récolte initiale des données conduit à un tableau brut de la forme

Individu	Variable
$w_1$	$X(w_1)$
$w_2$	$X(w_2)$
.	.
.	.
.	.
$w_k$	$X(w_k)$

Le nombre d'individus observé étant en général important, le tableau précédant ne permet pas d'analyser l'information obtenue. Il est donc nécessaire de créer un tableau plus synthétique où les observations identiques (possédant la même modalité) ont été regroupées.

Modalité	effectif
$x_1$	$n_1$
$x_2$	$n_2$
.	.
.	.
.	.
$x_k$	$n_k$

 **Definition: Effectif ou fréquence absolue**

On appelle effectif ou fréquence absolue de la modalité  $x_k$  le nombre  $n_k$  d'individus qui ont pris la modalité  $x_k$

 **Note**

$$n_1 + n_2 + \dots + n_k = N$$

## 2.2. Tableau de contingence.

Tableau de contingence : Ce tableau résume comment une caractéristique dépend d'une autre. Pour des raisons pratiques, on se limite généralement au tableau de contingence de deux variables qualitatives X et Y de modalités respectives  $(x_1, \dots, x_i, \dots, x_k)$  et  $(y_1, \dots, y_j, \dots, y_r)$ . Ce tableau donne le nombre d'individus possédant simultanément la modalité  $x_i$  de la variable X et la modalité  $y_j$  de la variable Y. Un tel tableau se présente sous la forme suivante, pour un échantillon de taille N.

Dans ce cas le tableau brut se présente sous la forme suivante:

Individu	Variable 1	Variable 2
$w_1$	$X(w_1)$	$Y(w_1)$
$w_2$	$X(w_2)$	$Y(w_2)$
.	.	.
.	.	.
.	.	.
$w_k$	$X(w_k)$	$Y(w_k)$

On désire créer un tableau appelé tableau de contingence donnant le nombre d'individus possédant simultanément la modalité  $i$  de variable 1 et la modalité  $j$  de variable 2 qui se présentera sous la forme suivante:

		Variable 2
		$y_1 \cdot \dots \cdot y_j \cdot \dots \cdot y_r$
Variable 1	$x_1$	$n_{11} \dots n_{1j} \dots n_{1r}$
	$\cdot$	$\dots\dots$
	$\cdot$	$\dots\dots$
	$x_i$	$n_{i1} \dots n_{ij} \dots n_{ir}$
	$\cdot$	$\dots\dots$
$\cdot$	$\dots\dots$	
$x_k$	$n_{k1} \dots n_{kj} \dots n_{kr}$	

où  $n_{ij}$  désigne le nombre de fois où X a pris la modalité  $x_i$  et Y la modalité  $y_j$ . Autrement dit,  $n_{ij}$  représente le nombre d'individus qui possède à la fois la caractéristique  $x_i$  et la caractéristique  $y_j$ .



## 2.3. Tableau des fréquences.

Variable discrète  
Variable continue.

9  
10

### 2.3.1. Variable discrète

On considère une variable quantitative discrète  $X$  à valeurs dans  $\{x_1, \dots, x_k\}$ . On suppose ses modalités ordonnées tels que  $x_1 < \dots < x_k$ . On commence par définir les quantités descriptives de bases.



#### Note

Pour une variable qualitative, les modalités ne sont pas mesurables. Pour une variable quantitative, les modalités sont mesurables. Ce sont

- des valeurs numériques ponctuelles lorsque la variable est discrète
- des intervalles lorsque la variable est continue ou lorsque la variable est discrète et qu'elle comporte beaucoup de modalités.



#### Definition: Fréquence relative

On appelle fréquence relative (ou simplement fréquence) de la modalité  $x_k$  le nombre  $f_k$  défini par  $f_k = n_k / N$ .



#### Definition: Fréquence relative cumulée

On appelle fréquence relative cumulée (ou simplement fréquence cumulée) de la modalité  $x_k$  le nombre

$F_k^{(c)}$  défini par :

$$F_k^{(c)} = f_1 + f_2 + \dots + f_k.$$



#### Note

$$f_1 + f_2 + \dots + f_k = 1.$$

On peut aussi définir les effectifs cumulés :



#### Definition: Effectifs cumulés

On appelle effectif cumulé de la modalité  $x_k$  le nombre  $E_k$  défini par :  $E_k = n_1 + n_2 + \dots + n_k$ .

La distribution des  $N$  observations de  $X$  peut être présentée sous la forme d'un tableau de fréquence où figurent, pour chaque modalité  $x_k$ , l'effectif  $n_k$ , la fréquence relative  $f_k$  et la fréquence cumulée  $F_k^{(c)}$  :

modalité	effectif	fréquence relative	fréquence cumulée
$x_1$	$n_1$	$f_1$	$F_1^{(c)}$
...	...	...	...
$x_i$	$n_i$	$f_i$	$F_i^{(c)}$
...	...	...	...
$x_k$	$n_k$	$f_k$	$F_k^{(c)}$

### Note

Les fréquences relative et cumulée peuvent être donnée sous forme de pourcentage.

### 2.3.2. Variable continue.

Dans le cas où la variable  $X$  est continue, la réalisation d'un tableau de fréquence nécessite au préalable une répartition en classes des données. On doit définir a priori le nombre de classes  $K$  et l'amplitude (ou l'étendue) de chaque classe. Ce choix doit résulter d'un compromis entre deux objectifs antagonistes : résumer les données ( $K$  ne doit pas être trop grand) sans perdre l'information pertinente ( $K$  ne doit pas être trop petit). Pour ce faire, un moyen «simple» est de diviser l'étendue des données en plusieurs intervalles de même longueur, puis l'on regroupe les classes d'effectifs trop petit (ie moins de 5 individus). On peut utiliser une des deux règles suivantes pour déterminer le nombre de classes :

Règle de Sturge:  $K=1+3/10\log_{10}(N)$ .

Règle de Yule:  $K=2,5^{0,25}$ .

L'intervalle entre les classes est alors donné par

Longueur de l'intervalle= $(x_{\max}-x_{\min})/k$ .

où  $x_{\max}$  (resp.  $x_{\min}$ ) désigne la plus grande (res. la plus petite) valeur de prise par les  $X(\omega)$ .

On note  $C_k$  l'ensemble des individus qui appartiennent à la classe  $K$ .

### Note

Le tableau regroupé en classe est souvent appelé distribution groupée. Si  $[a_i;a_{i+1}[$  désigne la classe  $j$ , on note, de manière générale :

$a_i$ :la borne inférieure de la classe  $i$ ,

$a_{i+1}$ :la borne supérieure de la classe  $i$ ,

$c_i=(a_i+a_{i+1})/2$  : le centre de la classe  $i$ ,

### Definition

L'amplitude  $L_k$  de la classe  $k$  est donnée par  $L_k = a_{i+1} - a_i$  c'est-à-dire la longueur de l'intervalle.

On peut alors définir la densité d'une classe :

### Definition

La densité  $d_k$  de la classe  $k$  est donnée par  $d_k = n_k / L_k$ .

Ce découpage en classes permet de se ramener au cas discret décrit précédemment pour obtenir le tableau de fréquences, en adaptant directement les définitions vues précédemment.

Tableau statistique pour une variable quantitative continue

Classes	Centres des classes $c_i$	Effectifs $n_i$	Fréquences $f_i$	Pourcentages $p_i$	Fréquences cumulées $F_i$
$[a_1; a_2[$	$c_1$	$n_1$	$f_1$	$p_1$	$F_1^{(c)}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$[a_i; a_{i+1}[$	$c_i$	$n_i$	$f_i$	$p_i = f_i * 100$	$F_i^{(c)}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$[a_k; a_{k+1}[$	$c_k$	$n_k$	$f_k$	$p_k$	$F_k^{(c)}$