

UNIVERSITE DE BLIDA 1  
FACULTÉ DES SCIENCES  
DEPARTEMENT DU TRONC COMMUN MATHEMATIQUES ET  
INFORMATIQUE(MI)

**Cours de Statistique descriptive à une seule variable**



**Réalisé par : Mme Messaoudi Nadia**

Chargée de cours : **1<sup>er</sup> année MI Section A**

Courriel : [messlina2012@gmail.com](mailto:messlina2012@gmail.com)

Année universitaire 2019- 2020

# Introduction

Ce polycopié de cours de statistique descriptive à une seule variable est destinée aux étudiants de première année du tronc commun de mathématique et informatique (MI) et conforme aux programmes du LMD. Ce cours présente des outils et méthodes de statistique. La statistique est une branche des mathématiques qui permet de recueillir, organiser, classer, résumer et présenter des informations statistiques qualitatives ou quantitatives en vue de les analyser pour en prendre des décisions judicieuses.

La statistique s'applique à la plupart des disciplines : agronomie, biologie, démographie, économie, sociologie, psychologie, médecine,...etc. On distingue deux principales branches des statistiques, la statistique descriptive et la statistique inférentielle. La statistique descriptive a pour but la recherche d'informations, la classification des observations qualitatives ou quantitatives, leur traitement pour les rendre utilisables et pour permettre leur interprétation. La statistique inférentielle (inductive) a pour objectif à partir d'hypothèses basées sur les probabilités, l'élaboration des modèles, l'établissement de lois et la définition de leurs propriétés.

Dans ce cours nous intéressons la statistique descriptive à une seule variable, dont on dispose des données qui peuvent soit concerner l'ensemble d'une population formée des individus, par exemple les étudiants de première année (MI), soit un échantillon, par exemple une section de 1<sup>er</sup> année (ce qui sera le cas le plus fréquent, étant donnée la difficulté à obtenir des données exhaustives sur l'ensemble d'une population : notamment en sciences humaines où la collecte de données peut faire intervenir un entretien avec chaque individu). A partir de ces données on peut faire l'objet d'une étude statistique, par exemple quel est l'âge moyen des étudiants de 1<sup>er</sup> année MI, ici l'âge est considéré comme étant une variable statistique. On pourra étudier d'autres propriétés sur l'ensemble des étudiants, par exemple : couleur des yeux, la série du bac, le sexe, la distance d'éloignement de leurs domicile à l'université, ...etc.

L'objectif du cours est de permettre aux étudiants de comprendre comment analyser, représenter les données et comment résumer leurs propriétés à l'aide des paramètres.

Le premier chapitre se focalisant sur le vocabulaire de la statistique descriptive univariée. Nous décrivons dans le chapitre deux les différents types de représentation graphiques des variables quantitatives ou qualitatives.

Nous présentons dans le troisième chapitre, quelques paramètres de tendance centrale et de dispersion d'une variable statistique. Nous fournissons également, autant d'exemples et de graphiques nécessaires afin d'avoir une meilleure compréhension et assimilation du cours.



# Chapitre 1

## Généralités sur la statistique descriptive

### 1.1 Terminologie

**Définition :** On définit les statistique (on dit une statistique) comme étant un ensemble d'informations qualitatives ou quantitatives sur un sujet précis.

**Exemple 1 :**

Une enquête dans un garage parking concernant la couleur des automobiles a donnée les résultats suivants :

Noire, verte, rouge, bleue, jaune, blanche, rouge, verte, blanche, bleue, verte, rouge, verte, rouge, verte, bleue. Le tableau suivant est un moyen synthétique de présentation des données :

Couleur	verte	rouge	bleue	blanche	noire	jaune
Nombre de voitures	5	4	3	2	1	1

**Exemple 2 :**

En vue d'étudier la note d'un module d'une section de 1<sup>er</sup> année d' étudiants du tronc commun qui est formée de sept groupes. On a questionné tous les étudiants de leurs notes. Les résultats ont été reportés dans le tableau suivant :

Note d'un examen	5	7	8	10	12	14
Nombre d'étudiants	5	20	13	40	12	10

**Exemple 3 :**

Dans le but d'étudier le nombre de pièces par logement, 100 logements ont été sélectionnés dans un quartier. On a dénombré le nombre de pièces qu'ils contient. Le tableau suivant a été obtenu :

Nombre de pièces	1	2	3	4	5	6
Nombre de logements	5	10	20	30	25	10

**Exemple 4 :** Afin d'étudier l'ancienneté du personnel cadre dans une entreprise de 50 employés, on a interrogé chacun des employés sur le nombre d'année écoulés dans l'entreprise. Les résultats de l'enquête sont donnés par le tableau suivant :

Durée en années	[0,4[	[4,8[	[8,12[	[12,16[	[16,20[
Nombre d'employés	8	22	8	7	5

### 1.1.1 Population

**Définition :** c'est un ensemble d'éléments homogènes, d'objets ou de personne auxquels on effectue une étude statistique.

*Exemple :*

Exemple $n^o$	Population
1	L'ensemble des 16 voitures dans un parking
2	L'ensemble des 100 étudiants de 1 <sup>er</sup> année tronc commun
3	L'ensemble des 100 logements
4	L'ensemble du personnel cadre d'une entreprise

### 1.1.2 Individu ou unité statistique

**Définition :** l'individu est un élément de la population.

**Exemple :** pour les exemples précédents l'individu est donné pour chaque exemple respectivement : l'étudiant, le véhicule, le logement et le travailleur.

### 1.1.3 Echantillon ou sous population

Souvent il est difficile d'observer toutes les données au lieu d'examiner l'ensemble total, on examine une partie ou un sous ensemble de la population qui est appelé *échantillon* ou *sous-population*.

Dans l'exemple 2 : les groupes 1, 2 et 3 d'une section de première année constituent un échantillon.

### 1.1.4 Caractère

Lorsque la population est proposée à une étude statistique c'est en fonction d'une propriété particulière que possèdent ses éléments. Un caractère est une propriété possédée par les individus permettant de les décrire et de les distinguer les uns des autres. Tout individu peut être décrit ou étudié selon un ou plusieurs caractères.

Un groupe d'étudiants, par exemple, peut faire l'objet d'une étude statistique concernant la taille ou le poids de ses éléments ou bien de leurs notes de module de mathématiques, leur âge, mention du leurs baccalauréat, l'éloignement de leur domicile par rapport à l'université,...etc. Les caractères présentés dans les exemples précédents sont données comme suit :

Exemple	Caractère
Exemple 1	Couleur des voitures
Exemple 2	Note de l'examen obtenue dans un module
Exemple 3	Nombre de pièces par logement
Exemple 4	Durée d'ancienneté des travailleurs

### Modalités d'un caractère

Les modalités d'un caractère se sont les diverses situations (ou cas) susceptibles d'être prises par le caractère. Un caractère peut posséder une ou plusieurs modalités.

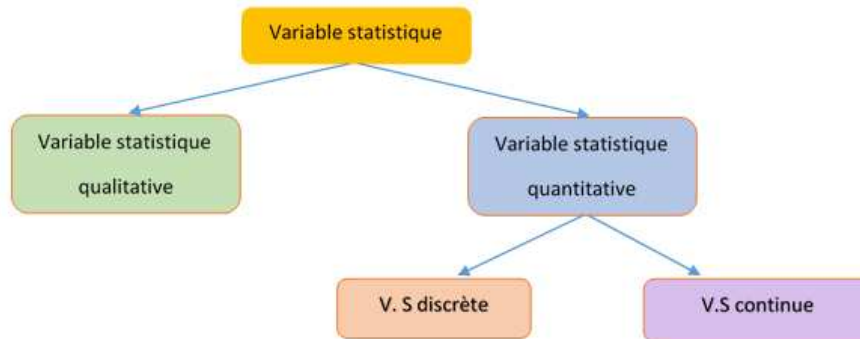
Le caractère mention, ses modalités sont : excellent ; très bien, bien , assez bien, passable. Le caractère âge des étudiants peut être décrit selon plusieurs modalités, il peut varier par exemple de 18 ans à 20 ans.

Dans l'exemple 1 ; le caractère couleur des voitures possède 6 modalités, qui sont : verte, rouge, bleue, blanche, noire et jaune.

Dans l'exemple 4, vous remarquez que les intervalles sont semi ouvert ( exemple :  $[0-4[$ , puis  $[4-8[$ ). Ceci pour éviter la possibilité qu'un individu ait deux modalités. La personne dont sa durée d'ancienneté est exactement 4 ans sera, classée uniquement dans le 2 ème intervalle.

## Propriétés du caractère

Un individu peut être décrit selon un caractère qualitatif ou quantitatif. On affecte à chaque modalité une variable statistique (V.S) qui peut être , quantitative ou qualitative. Le schéma suivant résume les différents type d'une V.S :



a/ *Le caractère est qualitatif* lorsque ses différentes modalités ne sont pas mesurables. Par exemple : on étudie un groupe de personne suivant la couleur des yeux, ou le groupe sanguins, ou l'état civil, la profession , ou bien le sexe, le diplôme obtenu, la mention, ou la nationalité.

b/ *Le caractère est quantitatif* :lorsque ses diverses modalités sont mesurables, à chacune des modalités on peut attacher un nombre ou une valeur. Ce nombre est appelé variable statistique. Celle ci peut être discrète ou continue.

b/1. Une variable statistique , notée  $x_i$  est dite *discrète* (ou discontinue) lorsqu'elle ne peut prendre que des valeurs isolées souvent des nombres entiers : 0,1,2,3,...  
Par Exemple ; le nombre de pièces par logements. La note d'un examen. Le nombre d'enfants des travailleurs.

b/2. Une variable statistique, notée  $x_i$  est considérée comme *continue* lorsque les diverses valeurs prises par elle sont en nombre indéterminée (des valeurs décimales) à priori dans un intervalle de valeurs.

Par exemple, l'âge d'un individu, n'est jamais déterminée avec exactitude.  
Dans l'exemple 3, la durée d'ancienneté des employés est une variable statistique continue.



### 1.1.5 Série statistique

**Définition :** la série statistique est la suite des observations (ou valeurs) prises par une variable  $x$  sur les  $N$  individus ( numérotés de 1 à  $N$ ). Les valeurs de la variable  $x$  sont notées :

$$x_1, x_2, \dots, x_N$$

#### Exemple

On s'intéresse à la variable *nombre d'enfants* noté  $x$  et à la série statistique des valeurs prises par  $x$  sur 12 personnes. Elle est donnée comme suit :

0	1	3	4	2	4	4	5	2	5	6	6
$x_1$	$x_2$										$x_{12}$

### 1.1.6 Distribution

**Définition :** une distribution est une série statistique ordonnée selon l'ordre croissant ou décroissant des observations.

Exemple : la série précédente est ordonnée comme suit : 0 ,1 ,2, 2, 3, 4, 4, 4, 5, 5, 5, 6, 6.

## 1.2 Tableau statistique

La représentation d'une population est composée de  $N$  individus décrites suivant un caractère à  $p$  modalités dans un tableau, consiste à dénombrer pour chacune des modalités  $i$  combien d'individus possèdent une des modalités puis les ordonner.

Modalité numéro $i$	1	2	...	$p$
Effectif $n_i$	$n_1$	$n_2$	...	$n_p$

- **Effectif partiel**  $n_i$  d'une modalité  $i$  est le nombre de fois que cette modalité apparaît.
- **Effectif total** de la population étudiée, notée  $N$  , la somme des  $p$  effectifs partiels  $n_i$  correspondant à chacune des valeurs, soit :

$$N = n_1 + n_2 + \dots + n_p = \sum_{i=1}^p n_i$$

Dans l'exemple 1,  $N = 16$  voitures. Ainsi, dans l'exemple 2, le nombre d'étudiants d'une section du 1<sup>er</sup> année est  $N = 100$  .

• **Fréquence relative** de la modalité  $i$ , notée  $f_i$  est un rapport des effectifs  $n_i$  correspondant à la modalité  $i$  sur l'effectif totale  $N$ . Donc :

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

Ce nombre est égal au proportion d'individus présentant la modalité  $i$  par rapport à l'ensemble de la population observée.

**Propriété :** La fréquence  $f_i$  est toujours comprise entre 0 et 1 , car :  $0 \leq n_i \leq N$

Pour une série statistique présentent  $p$  modalités, on a :

$$\sum_{i=1}^p f_i = f_1 + f_2 + \dots + f_p = 1$$

Preuve :

$$\sum_{i=1}^p f_i = \frac{n_1}{N} + \frac{n_2}{N} + \dots + \frac{n_p}{N} = \frac{n_1 + n_2 + \dots + n_p}{N} = 1$$

• **Fréquence en pourcentage :**  $f_i\% = f_i * 100$

**Exemple :** Reprenons le tableau de distribution de l'exemple 1.

Couleur $i$	verte	rouge	bleue	blanche	noire	jaune	Total
Effectif $n_i$	5	4	3	2	1	1	16
Fréquence $f_i$	5/16	4/16	3/16	2/16	1/16	1/16	1
$f_i\%$	31.25	25	18.75	12.5	6.25	6.25	100

Le nombre de voitures de couleur verte apparait 5 fois, d'où  $n_1 = 5$ .

Le pourcentage  $f_2\%$  ; signifie que 25% des automobiles sont de couleur rouge.

• **Effectif cumulé** croissant, noté  $n_{i\text{cum}} \nearrow$  ( respectivement décroissant  $n_{i\text{cum}} \searrow$ ) ou fréquence cumulée croissante  $f_{i\text{cum}} \nearrow$  ( respectivement décroissante  $f_{i\text{cum}} \searrow$ ) correspondant à la valeur  $i$  est le nombre d'individus (ou fréquence) ayant une valeur inférieure ou égale à  $i$  (respectivement supérieure ou égale à  $i$ ) , qui se calcul par :

$$n_{i\text{cum}} \nearrow = \sum_{k=1}^i n_k, i = 1, \dots, p$$

On a :

$$n_{i\text{cum}} \nearrow = n_1$$

$$n_p cum \nearrow = N$$

$$n_i cum \nearrow = n_{i-1} cum \nearrow + n_i, i = 2, \dots, p - 1$$

$$n_i cum \searrow = \sum_{k=i}^p n_k, i = 1, \dots, p$$

$$n_1 cum \searrow = N$$

$$n_p cum \searrow = n_p$$

De manière similaire pour les fréquences cumulées :

$$f_i cum \nearrow = \frac{n_i cum \nearrow}{N} = \sum_{k=1}^i f_k, i = 1, \dots, p$$

$$f_i cum \searrow = \frac{n_i cum \searrow}{N} = \sum_{k=i}^p f_k, i = 1, \dots, p$$

**Exemple :** Considérons l'exemple 3 du nombre de pièces qui apparaissent dans 50 logements.

Nombre de pièces	1	2	3	4	5	6	Total
Effectif $n_i$	5	10	20	30	25	10	100
Fréquence $f_i$	0.05	0.1	0.2	0.3	0.25	0.1	1
$F_i cum \nearrow$	0.05	0.05 + 0.1 = 0.15	0.15 + 0.2 = 0.35	0.35 + 0.3 = 0.65	0.65 + 0.25 = 0.9	0.9 + 0.1 = 1	
$F_i cum \searrow$	1	1 – 0.05 = 0.95	0.95 – 0.1 = 0.85	0.85 – 0.2 = 0.65	0.65 – 0.3 = 0.35	0.35 – 0.25 = 0.1	

On peut dire que 35% des logement ont moins que 3 pièces. Ainsi que, 85% des logements ont plus que 3 pièces.