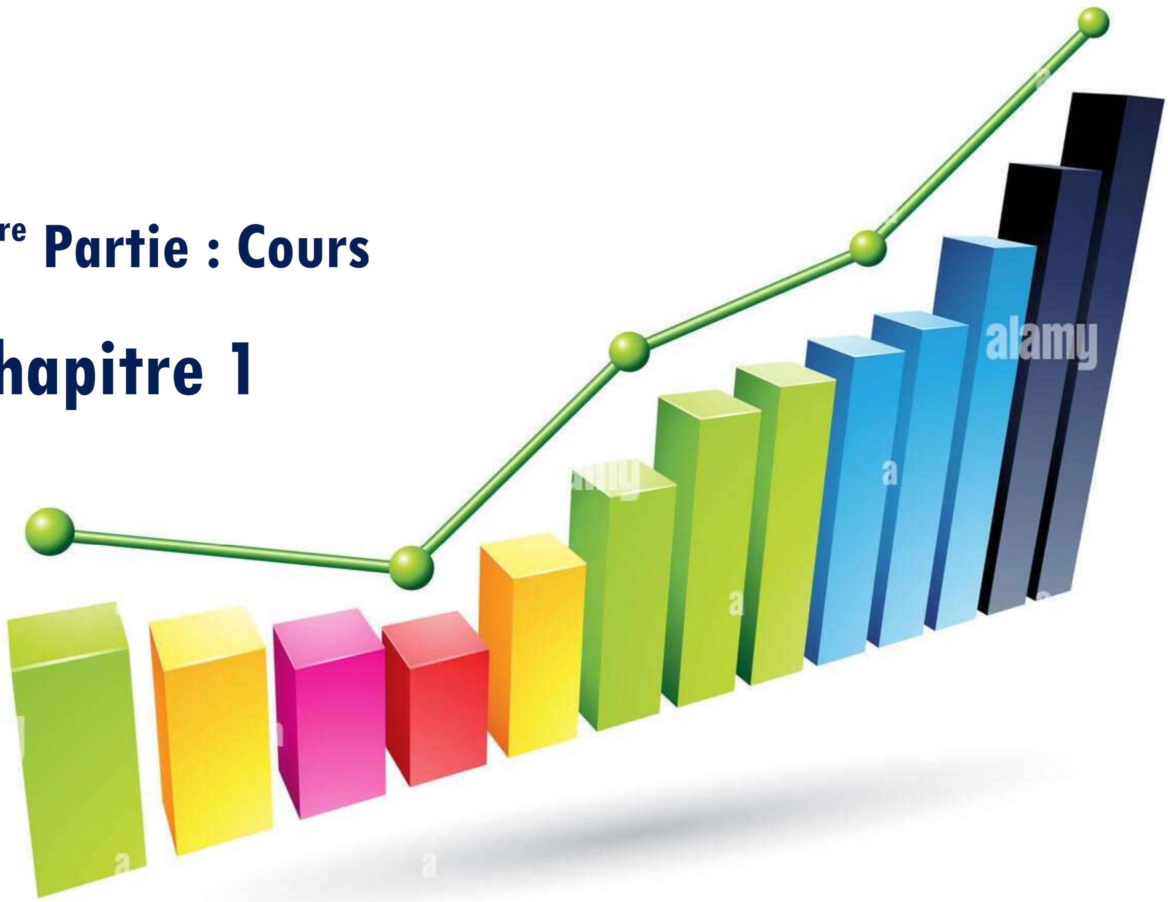
1ère Partie : Cours

Chapitre 1



Vocabulaire de la statistique descriptive

Ce chapitre est introductif, consacré à la définition de la statistique descriptive ainsi que des différents termes qui en constituent le vocabulaire de base.

En fait la statistique est une discipline qui concerne la quantification de certains phénomènes et l'élaboration de procédures et de règles pour étudier ces données quantitatives. Les méthodes statistiques donnent des procédures et des règles qui sont très utiles à l'analyse numérique.

L'intérêt qui est porté à cette discipline est démontré par son utilisation intense dans de nombreux domaines tels que les sciences expérimentales (biologie, physique, agriculture, agronomie, médecine, industrie,. . .), les sciences humaines, l'économie, etc.

D'après Bernard PY, dans son livre Statistique descriptive, comprendre et réussir (éditions Economica) : «La statistique descriptive est un ensemble de méthodes permettant de décrire et d'analyser, de façon quantifiée, des phénomènes repérés par des éléments nombreux, de même nature, susceptibles d'être dénombrés et classés».

1.1 Définitions fondamentales de la statistique

Pour un groupe d'individus ou d'objets la statistique est l'étude de :

1. La collecte de données.
2. Leur analyse, leur traitement et l'interprétation des résultats.
3. Leur présentation afin de rendre les données compréhensibles par tous.

Remarque 1.1.1 D'une manière générale, la méthode statistique est basée sur quatre concepts : la population, les variables (les caractères), les observations et les données.

1.1.1 Population statistique

Une population statistique est l'ensemble d'objets ou de personnes sur lequel on effectue des observations.

Exemples 1.1.1

1. Ensemble de personnes interrogées pour une enquête.
2. Ensemble de pays pour lesquels on dispose de données géographiques ou économiques.

1.1.2 Individus (ou unités statistiques)

Les individus sont les éléments de la population statistique étudiée. Pour chaque individu, on dispose d'une ou plusieurs observations.

Exemples 1.1.2

1. Chacune des personnes interrogées pour une enquête.
2. Chaque pays pour lequel on étudie des données socio-économiques, . . .
3. Chaque jour de l'année pour lequel on dispose de données météorologiques, ...

Remarque 1.1.2 Lorsqu'on observe qu'une partie de la population, on parle de sondage. La partie de la population étudiée est appelée échantillon et on cherche toujours à généraliserles résultats obtenus sur l'échantillon à toute la population.

Exemples 1.1.3

1. Etudier les notes du module d'analyse sur un échantillon de 200 étudiants est une expérience statistique. Les unités statistiques ou bien les individus correspondent aux étudiants. La population est l'ensemble des étudiants.

2. L'étude du chômage sur un échantillon des étudiants récemment diplômés du master 2, la population pourrait être l'ensemble des étudiants diplômés du master. Les unités statistiques ou bien les individus correspondent aux diplômés.

1.1.3 Caractère statistique (ou variable statistique)

Chaque individu d'une population est décrit par un ensemble de caractéristiques appelées variables ou caractères. On les notes souvent par des lettres majuscules : $X, Y \ldots$, donc c'est ce qui est observé ou mesuré sur les individus d'une population statistique.

Exemples 1.1.4

1. Chacune des personnes interrogées pour une enquête.
2. Chaque pays pour lequel on étudie des données socio-économiques, . . .
3. Chaque jour de l'année pour lequel on dispose de données météorologiques, ...

Remarque 1.1.3 Une variable doit donc présenter au minimum deux observations.

Exemples 1.1.5

1. La variables "sexe" a deux observations : masculin ou féminin.

2. Les observations de la variables "âge des ouvriers d'une entreprise", peuvent être : $[25, 30[\,;[30, 40[\,;[40, 45[$ et $[45, 60[$.

1.1.4 Modalité

Les modalités d'un caractère sont les différents résultats de l'observation (nombres ou propriétés).

Exemples 1.1.5

1. Cas qualitatif :

Les modalités de la variable $X =$ « situation familiale » sont :

$$M = \{\text{ célibataire, marié, veuf, divorcé }\}.$$

2. Cas quantitatif discret :

Les modalités du la variable $X =$ « Note à un examen » sont :

$$M = \{7, 9, 14, 16, 5\}.$$

3. Cas quantitatif continue :

Les modalités de la variable $X =$ « Taille » sont les valeurs appartenant aux intervalles $[150, 165[$, $[165, 180[$, etc. . .

Remarque 1.1.4 Les variables statistiques peuvent être classées selon leurs natures, d'où il y'a deux types variables qualitative ou quantitative.

1.2 Variable quantitative

Une variable statistique est quantitative si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

Remarque 1.2.1 Les variables quantitatives peuvent être discrètes ou continues.

1.2.1 Variable quantitative discrète

C'est une variable quantitative pouvant prendre par nature un nombre fini (ou dénombrable) de valeurs.

Exemples 1.2.1
1. Nombre d'enfants par famille.
2. Nombre de pièces d'un appartement.

1.2.2 Variable quantitative continue

C'est une variable quantitative pouvant prendre par nature une infinité de valeurs, généralement tout un intervalle réel.

Exemples 1.2.2

Tailles, poids, salaires, surfaces cultivées, température.

Remarque 1.2.2 Dans ce cas on utilise des intervalles $[a_i, b_i[$ au lieu de x_i.

1.3 Variable qualitative

Une variable statistique est qualitative si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage (ie une observation qui n'est pas mesurable).

Exemples 1.3.1
1. Sexe, situation familiale, . . .
2. Etat du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)

Remarque 1.3.1 Il y'a deux types de variables statistiques qualitatives : nominale ou ordinale.

1.3.1 Variable qualitative nominale

La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées (ne peuvent pas être classées).

Exemples 1.3.2

La variable X = « situation familiale » avec les modalités notées **c** (célibataire), **m** (marié), **v** (veuve), **d** (divorcé).

1.3.2 Variable qualitative ordinale

La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées.
Si $M = \{x_1, x_2, \ldots, x_r\}$ désigne l'ensemble des valeurs observées, ces valeurs sont ordonnées, c'est-à-dire :

$$x_1 \prec x_2 \prec \cdots \prec x_r.$$

La notation $x_1 \prec x_2$ se lit x_1 précède x_2.

Exemples 1.3.3

1. Un questionnaire de satisfaction demande aux consommateurs d'évaluer une prestation en cochant l'une des six catégories suivantes :
   (a) nulle, (b) médiocre, (c) moyenne, (d) assez bonne, (e) très bonne, (f) excellente.
   La variable X = « Niveau d'instruction ».

1.4 Effectif et fréquence d'une modalité (ou d'une classe)

1.4.1 Effectif

– L'effectif n_i d'une modalité (ou d'une classe) est le nombre de fois où la modalité (respectivement la classe) n° i a été observée.
– L'effectif total N est le nombre total d'individus observés.

$$N = n_1 + n_2 + \cdots + n_r = \sum_{i=1}^{r} n_i$$

1.4.2 Fréquence relative

– La fréquence (ou fréquence relative) f_i d'une modalité est le rapport de l'effectif n_i à l'effectif total N $\quad f_i = \dfrac{n_i}{N} = \dfrac{n_i}{\sum_{i=1}^{r} n_i}$

Remarque 1.4.1 Les fréquences relatives peuvent être exprimées en pourcentages, et on a le résultat suivant :

$$\sum_{i=1}^{r} f_i = 1$$

Car $\sum_{i=1}^{r} f_i = \sum_{i=1}^{r} \frac{n_i}{N} = \frac{1}{N}\sum_{i=1}^{r} n_i = \frac{N}{N} = 1$

Exemple 1.4.1

Sur 200 familles, 50 ont 2 enfants, on dira que la fréquence f_i correspondant à la valeur $x_i = 2$ de la variable « nombre d'enfants », est :

$$f_i = \frac{n_i}{N} = \frac{50}{200} = \frac{1}{4} = 0{,}25 \text{ soit } 25\%$$

1.5 Présentation dans un tableau statistique

1.5.1 Cas qualitatif nominale

Pour une variable statistique qualitative nominale, si l'ensemble $M = \{x_1, x_2, \dots, x_r\}$ désigne l'ensemble des modalités, alors le tableau statistique associé à ce caractère est

Modalités (numérotés) x_i	*Effectifs* n_i	*Fréquences* f_i
x_1	n_1	f_1
x_2	n_2	f_2
.	.	.
.	.	.
.	.	.
x_r	n_r	f_r
Total	N	1

Exemple 1.5.1

On s'intéresse aux valeurs de la variable X = «situation familiale» prises sur 20 personnes dont la codification est ;

c : célibataire, ***m*** : marié(e), ***v*** : veuf(ve), ***d*** : divorcé(e). Donc le domaine de la variable X est $M = \{x_1, x_2, \dots, x_r\}$.

Considérons les résultats suivants :

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

et on obtient le tableau suivant :

x_i	Effectifs n_i	Fréquences f_i
c	9	0,45
m	7	0,35
v	2	0,10
d	2	0,10
Total	20	1

Remarque 1.5.1 Avant d'aborder les autres cas on définit ce qui suit.

1.5.2 Fréquences cumulées croissantes F_i (f_{ic})

Pour $i = 1$, la fréquence cumulée croissante d'ordre 1 est $F_1 = f_1$.

Pour $i = 2$, la fréquence cumulée croissante d'ordre 2 est $F_2 = f_1 + f_2$

En général pour i quelconque la fréquence cumulée croissante d'ordre i est :

$$F_i = f_{ic} = f_1 + f_2 + \cdots + f_i = \sum_{p=1}^{i} f_p$$

1.5.3 Fréquences cumulées décroissantes F'_i (f_{id})

Pour $i = 1$, la fréquence cumulée croissante d'ordre 1 est $F'_r = f_r$

Pour $i = 2$, la fréquence cumulée croissante d'ordre 2 est $F'_{r-1} = f_r + f_{r-1}$.

En général pour i quelconque la fréquence cumulée décroissante d'ordre i est :

$$F'_i = f_{id} = f_r + f_{r-1} + \cdots + f_i = \sum_{p=i}^{r} f_p$$

Remarque 1.5.2 De la même manière on définit les effectifs cumulés croissants N_i (n_{ic}) et les effectifs cumulés décroissants N'_r (n_{id}).

$$N_i = n_{ic} = n_1 + n_2 + \cdots + n_i = \sum_{p=1}^{i} n_p$$

et

$$N'_i = n_{id} = n_r + n_{r-1} + \cdots + n_i = \sum_{p=i}^{r} n_p$$

Remarque 1.5.3 Les Fréquences cumulées et les effectifs cumulés sont utilisés pour les deux types de la variable quantitatif et la variable qualitatif ordinale.

1.5.4 Cas qualitatif ordinale et quantitatif discret

Si $M = \{x_1, x_2, \ldots, x_r\}$ désigne l'ensemble des modalités d'une variable qualitative ordinale, c'est à dire ces valeurs sont ordonnées : $x_1 \prec x_2 \prec \cdots \prec x_r$.
Avec $x_1 \prec x_2$ se lit x_1 précède x_2, ainsi le tableau associé est

x_i	*Effectifs* n_i	*Effectifs cumulés croissants* N_i	*Fréquences* f_i	*Fréquences cumulées croissantes* F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_r	n_r	$N_r = N$	f_r	$F_r = 1$
Total	N		1	

Exemple 1.5.2
20 chemises sont classées par taille :
$x_1 = \mathrm{S}$, $x_2 = \mathrm{M}$, $x_3 = \mathrm{L}$, $x_4 = \mathrm{XL}$ et $x_5 = \mathrm{XXL}$. Le tableau associé est

x_i	*Effectifs* n_i	*Effectifs cumulés croissants* N_i	*Fréquences* f_i	*Fréquences cumulées croissantes* F_i
x_1	4	4	0,20	0,20
x_2	2	6	0,10	0,30
x_3	5	11	0,25	0,55
x_4	8	19	0,40	0,90
x_5	1	20	0,05	1,00
Total	20		1,00	

Remarque 1.5.4 Le cas quantitatif discret se fait de la même manière que le cas qualitatif ordinal, et on obtient un tableau statistique semblable à celui du cas qualitatif ordinal.

1.5.5 Cas quantitatif continue

Dans le cas quantitatif continu on aura le tableau suivant

Classes $[b_{i-1}, b_i[$	*Centres* c_i	n_i	N_i	f_i	F_i
$[b_0, b_1[$	c_1	n_1	N_1	f_1	F_1
$[b_1, b_2[$	c_2	n_2	N_2	f_2	F_2
.	.	.	.	.	.
.	.	.	.	.	.
$[b_{r-1}, b_r[$	c_r	n_r	$N_r = N$	f_r	$F_r = 1$
Total		N		1	