

CHAPI

Notions de base et vocabulaire statistique.

(1)

I.1) Introduction:

La statistique est une ^{méthode scientifique qui}
(outil)
consiste à réunir des données chiffrées sur des ensembles
nombreux, puis à analyser, à commenter et à critiquer
ces données.

On observe une différence entre la statistique qui est un
outil scientifique et une statistique qui est un
ensemble de données chiffrées sur un sujet précis.
La statistique descriptive est une méthode scientifique
qui permet de recueillir, classer, présenter et analyser
des informations statistiques provenant des ensembles
comportant un grand nombre d'éléments.
Elle est utilisée dans de nombreux domaines tels que
la biologie, physique, agronomie, l'économie... etc.

I.2) Concepts de base de la statistique.

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

un individu et tout élément de la population Ω .
il est noté $w \in \Omega$
l'individu ou l'unité statistique et un objet pour
lequel nous sommes intéressés à recueillir des
informations.

exemple:

Si on étudie la production annuelle d'une usine
de boîtes de boisson en métal (canettes).

La population Ω est l'ensemble des boîtes produites
durant l'année et une boîte constitue un individu.

I.2.1: Population - individu et échantillon (2)

On appelle population l'ensemble ^{d'objets} sur lequel porte l'étude statistique. Il est noté généralement par N . Une population peut être de nature très variée (êtres humains, animaux, plantes, bactéries, ... etc). Les objets sont appelés individus ou (unités statistiques). ^{← vers} Lorsqu'on n'observe qu'une partie de la population, on parle de sondage. La partie de la population étudiée est appelée échantillon et on cherche toujours à généraliser les résultats obtenus sur l'échantillon à toute la population.

Exemples:

- 1) Étudier les notes du module de statistique sur un échantillon de 1000 étudiants est une expérience statistique.
Les individus correspondent aux étudiants.
La population est l'ensemble des étudiants.
- 2) Si l'on s'intéresse maintenant à la circulation automobile dans une ville, la population est alors constituée de l'ensemble des véhicules susceptibles de circuler dans cette ville à une date donnée. Dans ce cas, N = ensemble des véhicules.

I.2.2 Variable statistique (ou caractère) (3)
chaque individu d'une population et de lui par un
ensemble de caractéristiques appelés variables ou
caractères. On le représente souvent par des lettres
majuscules $X, Y, \dots$ etc.

Les valeurs que peut prendre une variable statistique
sont appelés modalités.

Une variable doit donc présenter au minimum
deux modalités.

exple de variables: la couleur, le sexe, le poids,
la taille, la marque, le modèle, le prix, ... etc.

* La variable "sexe" a 2 modalités: masculin ☐ ou
féminin ☐

* Les modalités de la variable "âge des ouvriers d'une
entreprise" peuvent être $[25, 30[$; $[30, 40[$; $[40, 45[$; $[45, 60[$.

Remarque: Les variables statistiques peuvent être
classées selon leur nature en deux catégories.

a) Variable quantitative:

Une variable est dite "quantitative" si ses différentes
valeurs sont mesurables.

— Il y a deux types de variables quantitatives

— Variable quantitative discrète:

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

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

— Variable quantitative continue:

— Variable quantitative continue:
Une variable quantitative est dite continue si elle prend toutes ses valeurs dans des intervalles de l'ensemble des nombre réels;
Ex. "taille en cm" observée sur un échantillon de personnes. C'est une variable continue.

prend toutes les valeurs de
l'ensemble des nombres réels;
exple: La variable "taille en cm" observée sur un
échantillon de 500 individus est une variable
quantitative continue, dont les résultats sont le suivants:
100 individus ont leur taille comprise dans l'intervalle [150, 160[
250 = " " " " " "[160, 170[
et 150 dans l'intervalle [170, 185].
La variable "couleur des yeux" est une variable
qualitative; elle ne prend que quelques-unes de ses différentes
valeurs possibles.

100 millions
250 =
et 150 dans l'intervalle $[170, 185[$
10 qualitative; qualitative

b) Variable qualitative ;
variable est dite qualitative

250
et 150 dans l'intervalle $[150, 250]$

b) Variable qualitative;
Une variable est dite qualitative si ses différentes modalités ne sont pas mesurables et si à dire elle ne peuvent pas être ordonnées (si on considère 2 modalités de ce caractère prises au hasard, on ne peut pas dire si l'une est inférieure ou égale à l'autre).

exemple; le sexe; la nationalité; la couleur des yeux, ... etc. { 5

I.2.3) observations;

Une variable donne lieu à plusieurs observations.

Si la variable est notée X , alors on notera par $x_1, x_2, x_3, \dots, x_n$ les observations et on constitue ainsi

une série statistique où n représente le nombre d'observa^{ons}

Si les observations portent sur la totalité de la population, on dit que n est la taille de la population ou effectif total

Si les observations n'ont porté que sur une partie de la population, on dit que n est la taille de l'échantillon de la population observée.

exemple; Soit un échantillon de 10 individus, la variable âge observée peut prendre les valeurs suivantes;

$x_1 = 31; x_2 = 40; x_3 = 25; x_4 = 30; x_5 = 35; x_6 = 40; x_7 = 28;$
 $x_8 = 42; x_9 = 33; x_{10} = 37$

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

I.2.4) Données

Lorsque les observations portent sur une ou plusieurs variables (on note p variables) et sur un ou plusieurs individus (on note n individus),

l'ensemble des valeurs des observations $\{6$ récoltées constituent les données de l'étude ou de l'expérience effectuée. On peut les mettre sous forme d'un tableau (individu/variable) à n lignes et p colonnes.

variables Individus	X_1	X_2	...	X_p
I_1	X_{11}	X_{12}	...	X_{1p}
I_2	X_{21}	X_{22}	...	X_{2p}
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
I_n	X_{n1}	X_{n2}	...	X_{np}

chaque individu I est représenté par un numéro d'ordre $i=1,2,\dots,n$ et chaque variable X est indexée d'un numéro d'ordre correspondant $j=1,2,\dots,p$.
 X_{ij} est la valeur prise par la $j^{\text{ème}}$ variable pour le $i^{\text{ème}}$ individu.

exple: Soit un échantillon de 5 individus sur lesquels on observe 2 variables X_1 le poids et X_2 "l'âge". Les résultats de cette étude sont représentés dans le tableau suivant.

	$X_1 = \text{poids}$	$X_2 = \text{âge}$
I_1	40	12
I_2	50	18
I_3	65	25
I_4	75	30
I_5	80	42

I.3) Tableaux statistiques

(7)

I.3.1) Présentations des données sous forme de Tableaux

Au cours d'une étude statistique, les données sont recueillies de façon désordonnée. Cette masse d'informations ne peut fournir de renseignements que si l'information en question est classée et mise en ordre.

Le classement ne peut se faire que dans des tableaux statistiques qui servent de documentation statistique. Un tableau statistique est constitué des modalités x_i (nbe d'observations) et des effectifs n_i .

- Les tableaux statistiques doivent comporter :
- * Un titre (indiquant l'objet du tableau).
 - * L'unité utilisée (dinars, kg, km, milliers de tonnes, millions, ...etc)
 - * Les titres de lignes et de colonnes.
 - * Les sources qui doivent être citées.

Remarque 2:

- * La taille de la population étudiée (échantillon) s'appelle effectif total noté N . Il est égal à la somme des effectifs partiels n_i ($N = \sum_{i=1}^k n_i$) appelés aussi fréquences absolues n_i (k est le nbe de modalités) ou

* Le rapport $f_i = \frac{n_i}{N}$ est appelé fréquence ou (8)

fréquence relative. (elle représente la proportion d'individus présentant la modalité x_i).

* On peut remplacer f_i par $f_i \times 100$ qui représente alors un pourcentage.

* La somme des fréquences relatives est égale à 1.

$$\left(\sum_{i=1}^k f_i = 1 \right)$$

La présentation des tableaux statistiques }
diffère selon la nature de variable

A) Cas d'une variable qualitative.

Lorsque le caractère étudié est qualitatif, x_i va représenter les différentes modalités du caractère. (qu'on note M_i)

et n_i ou f_i ses fréquences.
Le tableau statistique d'une telle variable se présente alors sous la forme suivante:

Modalités M_i du caractère	M_1	M_2	...	M_k	Total
Nombre d'individus n_i (effectif)	n_1	n_2	...	n_k	N

ou $\hat{M}$ chose en colonne ou en ligne.

B) Cas d'une variable quantitative discrète (9)

Si on note $x_1, \dots, x_k$ les différentes valeurs (modalités) d'un caractère quantitatif discret et les effectifs partiels respectifs $n_1, \dots, n_k$. Son tableau statistique se présente alors sous la forme suivante:

valeurs x_i	x_1	x_2	...	x_k
Effectifs n_i	n_1	n_2	...	n_k

exple ①: Répartition des logements selon le nombre de pièces

Nombre de pièces x_i	1	2	3	4	5	6	Total
Effectifs n_i	5	10	20	30	25	10	100

exple ②: Une étude sur la répartition des ménages selon le nombre de personnes qui les composent nous donne le tableau suivant;

Nombre de personnes dans un ménage x_i	Nombre de ménages n_i
1	50
2	80
3	90
4	120
5	350
6	160
7 et plus	150
Total	1000

plus de 10
personnes

exemple:

Une enquête sur la destination de 12 touristes algériens, a permis d'avoir les résultats suivants;

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

La variable étudiée représente "les destinations des touristes algériens".

Elle présente 5 modalités:

Tunisie, France, Turquie, Maroc, Espagne.

avec les effectifs partiels respectifs

$n_1 = 4$; $n_2 = 3$; $n_3 = 3$; $n_4 = 1$; $n_5 = 1$

de là, on obtient le tableau statistique correspondant

Destination	Effectif partiel n_i	Fréquence f_i
Tunisie.	4	0,333 ^{0,3333...}
France.	3	0,25
Turquie	3	0,25
Maroc	1	0,083 ^{0,0833...}
Espagne	1	0,083 ^{0,0833...}
Total	12	0,999 ≈ 1

remarque 3: Dans cet exemple, la Σ des fréquences est proche de 1 et ceci est dû aux arrondis des fréquences (trop de calculs après les virgules; erreurs d'arrondi)

C) Cas d'une variable quantitative continue

180

Lorsque la variable est continue, on regroupe les observations en k -classes d'extrémités $c_0, c_1, \dots, c_k$. (ces classes sont disjointes et leur réunion recouvre la totalité des valeurs possibles).

Ces classes sont notées $[c_i, c_{i+1}[$.

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

Le nombre n_i est l'effectif de la $i^{\text{ème}}$ classe.

On appelle "centre" d'une classe $[c_i, c_{i+1}[$ la valeur $y_i = \frac{c_i + c_{i+1}}{2}$ et "l'amplitude" de cette classe est notée $a_i = c_{i+1} - c_i$.

D'une manière générale, une série statistique d'une variable quantitative continue peut se représenter à l'aide d'un tableau du type suivant:

classes	$[c_1, c_2[$	$[c_2, c_3[$	...	$[c_k, c_{k+1}[$
Effectifs n_i	n_1	n_2	...	n_k

Exemple: On a observé les tailles (en cm) de 24 personnes

les résultats sont les suivants:

170; 165; 170; 165; 175; 165; 178; 178; 162; 159;
 158; 176; 180; 180; 182; 179; 183; 165; 176; 177;
 180; 173; 173; 162.

Les observations peuvent être représentées par des classes d'amplitude constantes égales à 4 cm par ex. (11)

Taille	Effectifs n_i	Fréquences f_i	Fréquences en pourcentage
[158, 162[	2 = n_1	$f_1 = \frac{n_1}{N} = \frac{2}{24} \approx 0,083$	8,3 %
[162, 166[	6 = n_2	0,25	25 %
[166, 170[	0 = n_3	0	0 %
[170, 174[	4 = n_4	0,167 0,166666667	16,7 %
[174, 178[	4 = n_5	0,167	16,7 %
[178, 182[	6 = n_6	0,25	25 %
[182, 186[	2 = n_7	0,083	8,3 %
total	24 = N	1	100 %

Règle de STURGES

12

Cette règle est utilisée pour déterminer le nombre de classes à utiliser pour représenter une variable statistique quantitative continue. Le nombre de classes est égal à $\lceil 1 + 3,3 \log n \rceil$ le plus proche de la quantité $1 + 3,3 \log n$.

Exemple: Pour un échantillon de taille $n = 200$, on doit utiliser $1 + 3,3 \log 200 \approx 8,59 \approx 9$ classes.

L'amplitude constante de ces classes sera égale à :

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

I.4) Représentations Graphiques

I.4.1) Introduction

Un tableau statistique contient toute l'information prélevée sur l'échantillon. mais il est très utile de traduire ces informations par un graphique. afin d'avoir une synthèse visuelle. Suivant la nature de la variable étudiée, on utilise différents types de représentations.

I.4.2) Cas de variables qualitatives

13

Une variable qualitative peut être représentée graphiquement de trois manières.

① Représentation circulaire par des secteurs
Elle consiste à représenter sur un cercle chaque modalité par un secteur dont le degré est déterminé à l'aide de la règle de trois de la manière suivante.
Supposons qu'on veuille calculer le degré du secteur représentant la modalité " i " d'effectif partiel n_i , alors on peut écrire :

$$\begin{array}{ccc} N \text{ (effectif total)} & \longrightarrow & 360^\circ \\ n_i & \longrightarrow & d_i = \frac{n_i}{N} \times 360^\circ = f_i \times 360^\circ \end{array}$$

$$\forall i = 1, \dots, k.$$

exemple: on a étudié le régime des études sur un échantillon de 200 élèves d'un lycée donné, les résultats obtenus sont les suivants;

Régime des études	Extérieure	Intérieure	Demi-pensionnaire	Total
Nombre d'étudiants n_i	70	50	80	200
Fréquences f_i	0,35	0,25	0,4	1

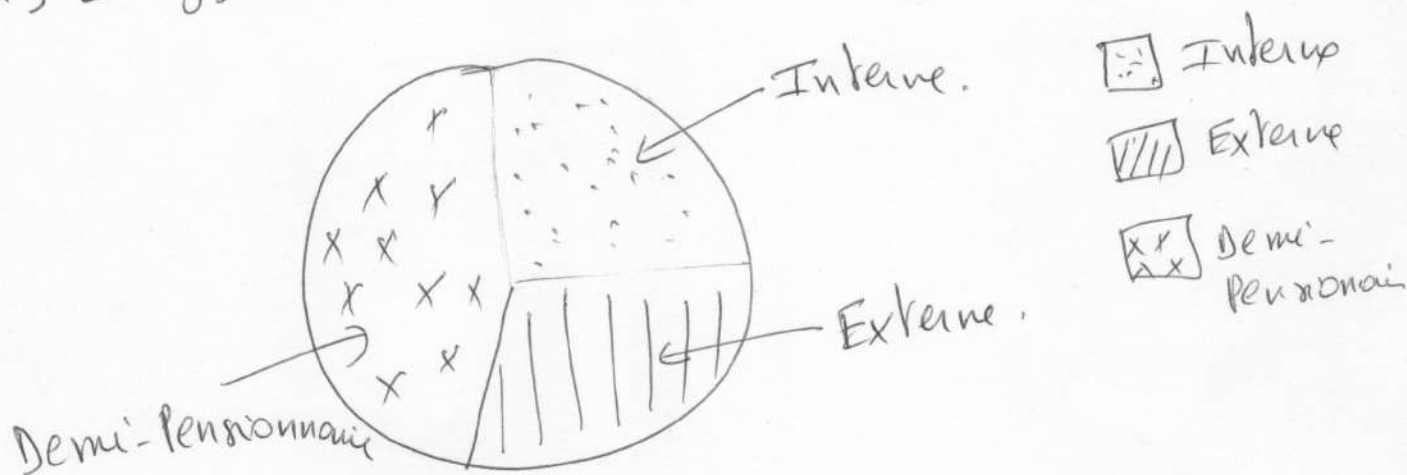
Donner la représentation circulaire de ce tableau statistique.
(traduire ces résultats par une représentation circulaire)

Calculons le degré correspondant aux différents secteurs: 14

$$d_1 = f_1 \times 360^\circ = 0,35 \times 360^\circ = 126^\circ$$

$$d_2 = f_2 \times 360^\circ = 0,25 \times 360^\circ = 90^\circ$$

$$d_3 = f_3 \times 360^\circ = 0,4 \times 360^\circ = 144^\circ$$



B) Représentation en rayons d'orgue

Elle consiste à porter en abscisse les modalités de la variable étudiée, au dessus desquelles on trace des rayons d'orgue de même largeur et en ordonnée les effectifs ou les fréquences qui vont déterminer leur hauteur.

exemple: Une enquête auprès de 500 étudiants de 1^{ère} année admis en 2^{ème} année, sur les spécialités préférées a fourni les résultats suivants;

Spécialité préférée	Nombre d'étudiants n _i
Mathématiques (M)	40
Physique (P)	80
Chimie (C)	50
Informatique (I)	150
Recherche Opérationnelle (R.O)	80
T.C.T	100

Les résultats peuvent être traduits par le diagramme (15) en barreaux d'orgue suivant:

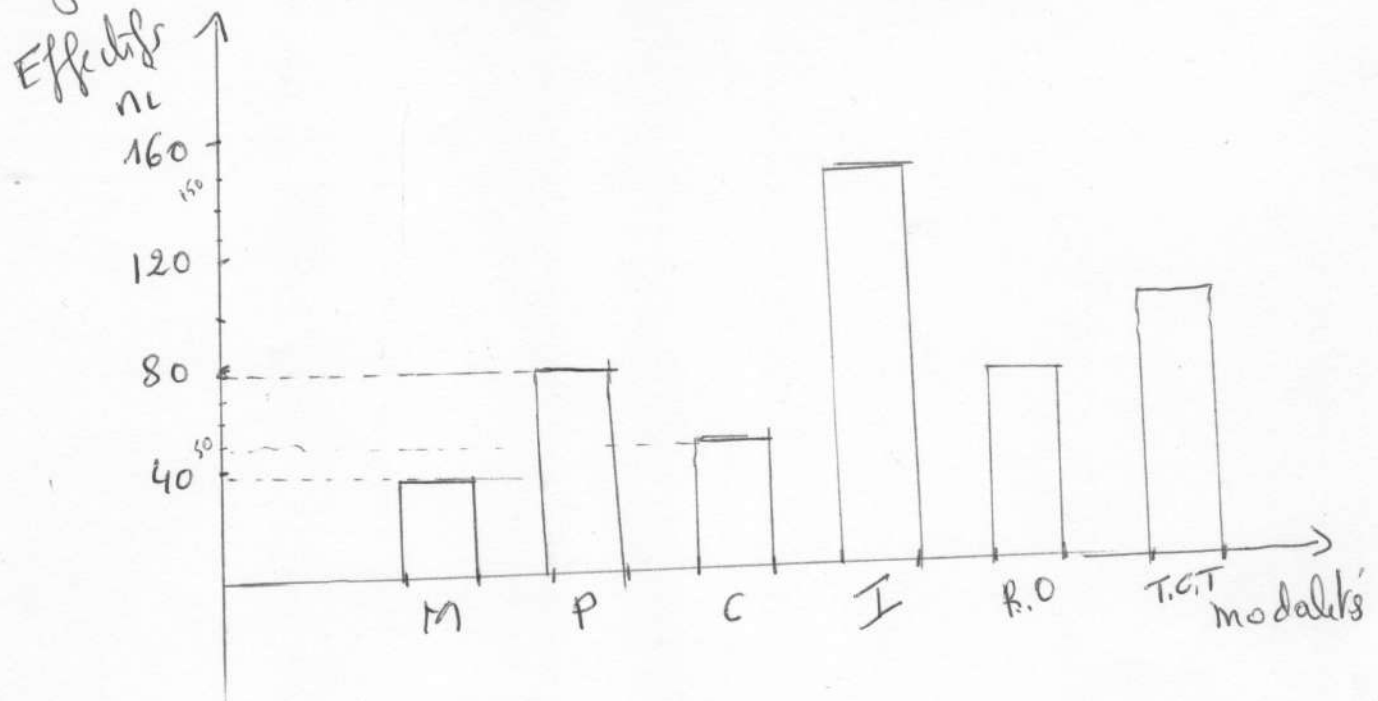


Diagramme en barreaux d'orgue.

(C) Diagramme en bandes:

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

exple: Les ventes réalisées par une maison de fabrication automobile durant les années 2000 à 2004 sont les suivantes:

Véhicules	2000		2004	
	n_i	f_i	n_i	f_i
Deux portes	800	0,16	1600	0,23
Petites 4 places	1500	0,3	2000	0,28
Cinq places	1700	0,34	900	0,13
Modèle luxe	1000	0,2	2500	0,36
Total.	5000	1	7000	1

Le diagramme en bandes correspondant à cette variable est le suivant:

(16)

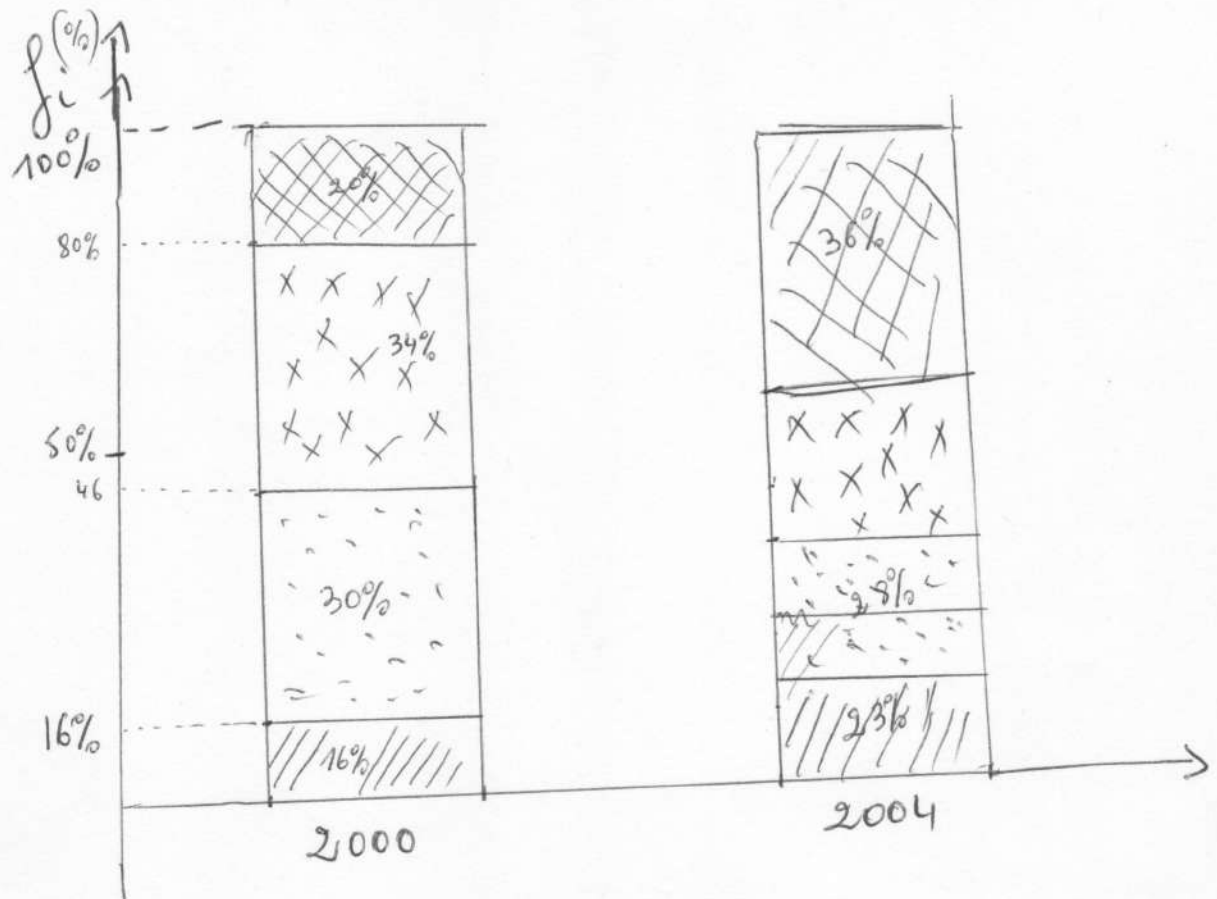
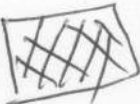


Diagramme en bandes

 modèle luxe.

 cinq places

 petites 4 places

 Deux portes.

I.4.3) Cas des variables quantitatives

(17)

Nous allons distinguer les diagrammes correspondant à une variable discrète et ceux correspondant à une variable continue.

A) Diagramme en bâtons

Il consiste à représenter les diverses valeurs x_i prises par une variable discrète, en fonction des fréquences f_i ou des effectifs n_i de la manière suivante :
On représente en abscisse les observations x_i et au dessus de chaque valeur, on trace un bâton dont la hauteur est proportionnelle à f_i ou à n_i .

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

x_i	55	60	65	70	75	80	Total (Σ)
n_i	20	60	40	20	50	10	200
f_i	0,1	0,3	0,2	0,1	0,25	0,05	1

Cette variable est discrète donc on peut la représenter graphiquement par le diagramme en bâtons en utilisant les fréquences f_i (ou) les effectifs n_i .

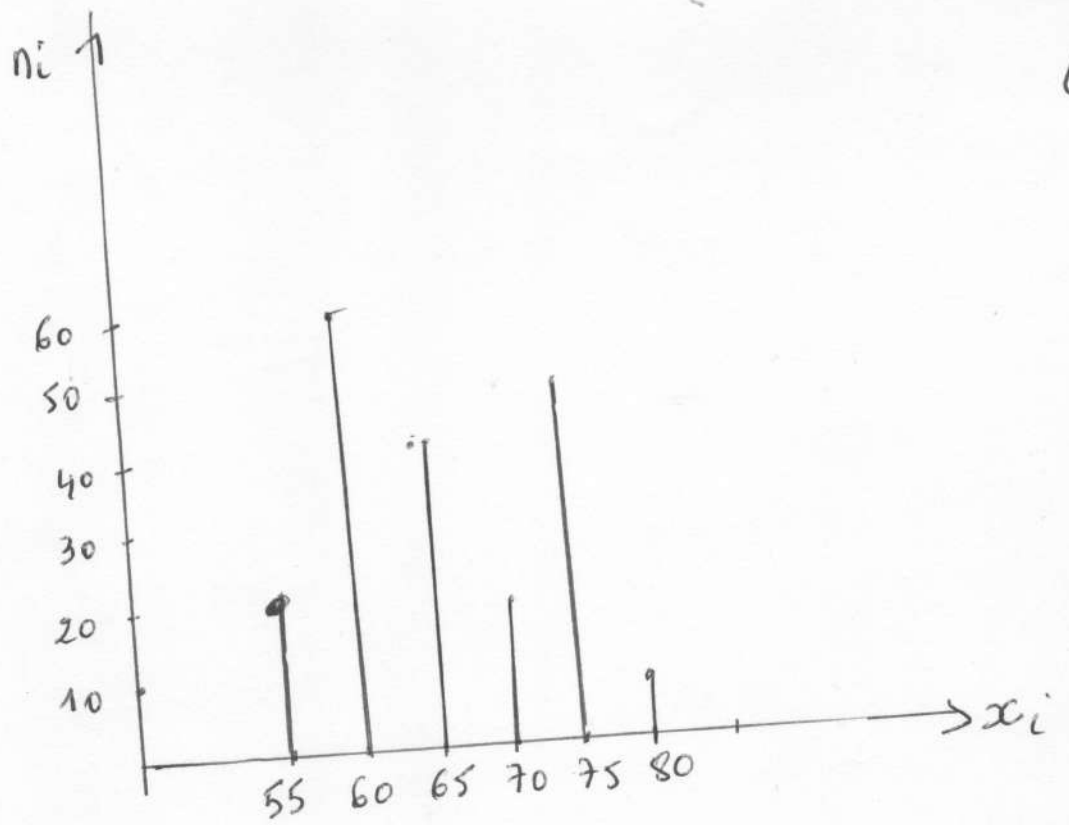


Diagramme en bâtons

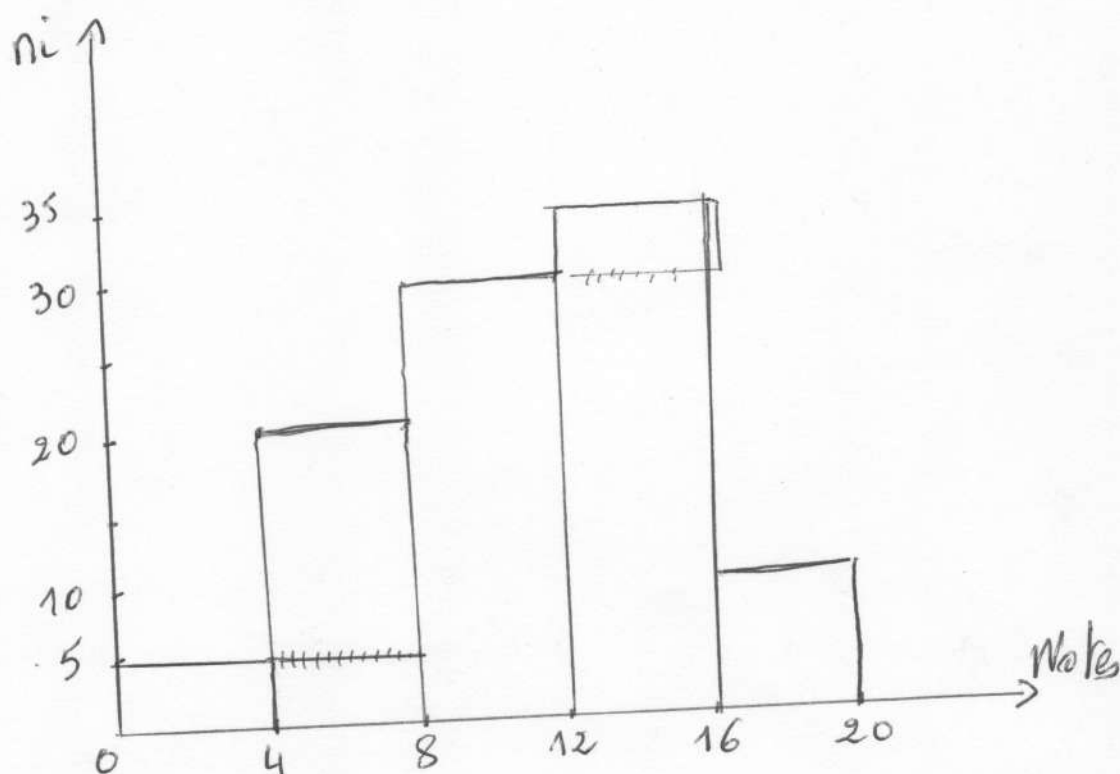
B) Histogramme

L'histogramme est utilisé pour représenter une variable continue. On représente sur l'axe des abscisses les différentes classes (supposées de même amplitude) et au-dessus de chacune d'elles on trace des rectangles dont les hauteurs sont directement proportionnelles aux fréquences ou aux effectifs.

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

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

La variable étant continue dont elle peut être représentée graphiquement au moyen de l'histogramme suivant:



Histogramme

Remarque 4:

Le mode de représentation est valable seulement dans le cas où on a des classes d'intervalles d'amplitude constante. Dans le cas où les amplitudes sont différentes, pour tracer l'histogramme, on utilise la méthode des effectifs rectifiés qui consiste à multiplier l'effectif de chaque classe par

le rapport entre l'amplitude de la plus petite classe sur l'amplitude de la classe considérée, on obtient ainsi les effectifs rectifiés notés n'_i avec

(20)

$$n'_i = n_i \times \frac{\text{la plus petite amplitude}}{\text{amplitude de la classe } i}$$

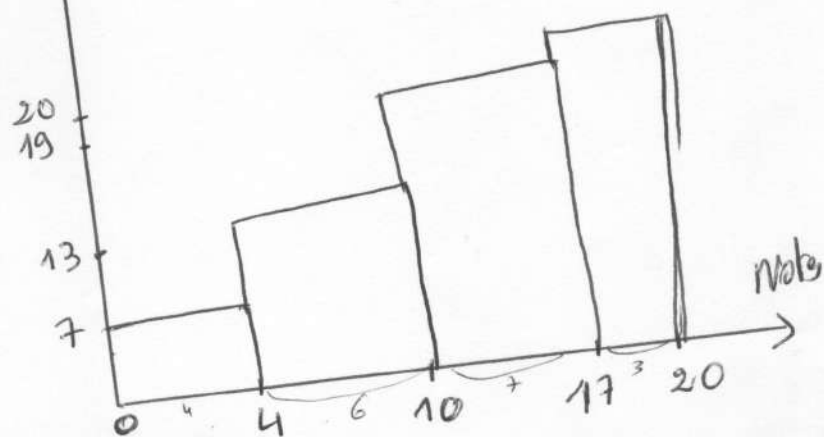
Exemple: On reprend l'exple précédent avec une nouvelle distribution.

Notes	$[0, 4[$	$[4, 10[$	$[10, 17[$	$[17, 20]$	Total
Effectif n_i	10	25	45	20	100
Amplitude a_i	4	6	7	3	
Effectif rectifiés n'_i	$10 \times \frac{3}{4} \approx 7$	$25 \times \frac{3}{6} \approx 13$	$45 \times \frac{3}{7} \approx 19$	$20 \times \frac{3}{3} = 20$	

longueur de la classe

$\bar{a}$ ajout

d'où l'histogramme correspondant.



Remarque 5:

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

(21)

Exemple:

Notes	$[0, 4[$	$[4, 10[$	$[10, 17[$	$[17, 20]$	Total.
n_i	10	25	45	20	100
f_i	0,1	0,25	0,45	0,20	1
a_i	4	6	7	3	—
$\frac{f_i}{a_i}$	0,025	0,042	0,064	0,066	—

$\frac{f_i}{a_i}$

0,066
0,064

0,042

0,025

0 4 10 17 20 Notes

Histogramme.

Remarque 6: La surface de chaque rectangle est égale à la fréquence de la classe correspondante ($S_i = \frac{f_i}{a_i} \times a_i = f_i$) donc la surface totale de l'histogramme est égale à 1.

C.4.4) Polygone

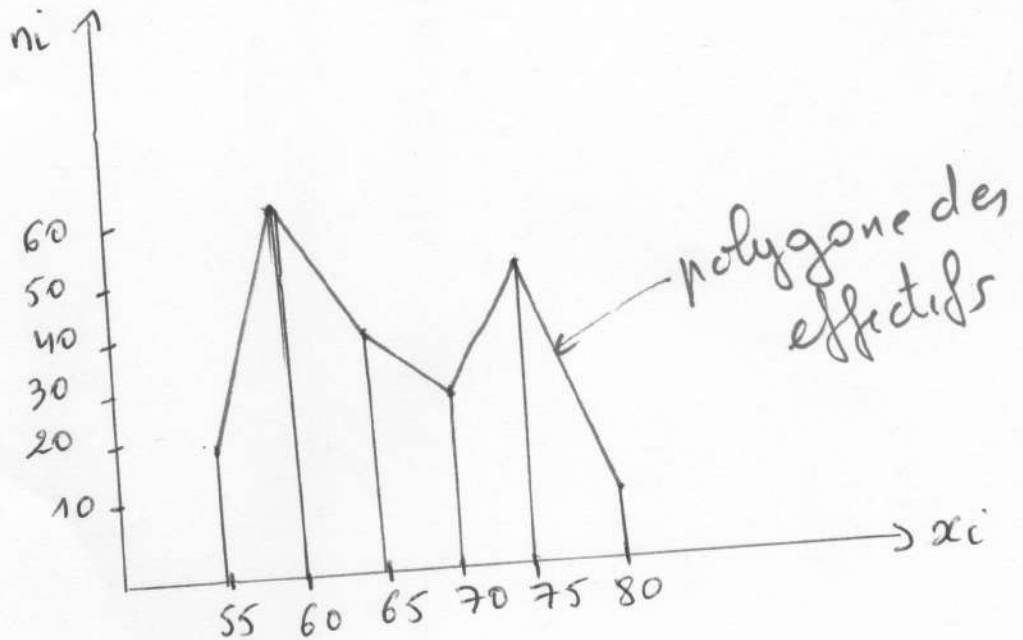
(22)

Le polygone est construit à partir du diagramme en bâtons

ou de l'histogramme.

Il s'agit d'une ligne brisée qui joint les points supérieurs des bâtons ou les points milieux des côtés supérieurs des rectangles successifs de l'histogramme.

* Le polygone des effectifs correspondant à la variable étudiée dans l'exemple (A) est:



* Le polygone des effectifs correspondant à la variable étudiée dans l'exemple B est:

