Université Abderrahmane Mira de Bejaia
Faculté des sciences humaines et sociales
Département des sciences sociales

Intitulé du module

Statistique descriptive

Destiné aux étudiants de la première année

Semestre 1

Cours préparés par

Dr AMOUR-Mustapha

q

Année Universitaire : 2021-2022

Généralités et définitions de la statistique descriptive

Cours 1:

Introduction:

Pour avoir une idée de la place et l´importance que revêt la statistique dans la vie de tous les jours, il suffit de parcourir les différents journaux ou revues qui proposent des résultats observés á l´aide d´études ou de sondage. Nous infiltre sans avoir á manipuler des masses plus au moins considérables de données.

A. Définition, nature et objet de la statistique:

Le mot "statistiques" au pluriel, désigne des collections de chiffres, présentés sous forme de tableaux parfois sous forme de graphiques, qui regroupent toutes les observations effectuées sur les faits nombreux, relatifs à un même phénomène.

Il est bien que qu'une pareille définition ne saurait suffire. Le statisticien, placé en face ces relevés chiffrés; devra les examiner de plus près," de manière à obtenir des supports numériques sensiblement indépendants des anomalies du hasard et qui dénotent l'existence de causes régulière dont l'action est combinée avec celle des causes fortuites".

La statistique au singulier, c'est l'ensemble des procédés ou des méthodes qui auraient pour but l'étude mathématiques des statistiques.

Si nous retenons ces deux définitions, du terme "statistique", nous dirons que la statistique est l'outil avec lequel on travaille une matière première constituées par les statistiques.

La statistique est une branche des mathématiques appliquées dont les principes découlent de la théorie des probabilités et qui a pour objet le groupement méthodique ainsi que l'étude des séries de faits ou de données numériques.

La statistique est la science qui a pour objet de recueillir un ensemble de données numériques relatives à tel ou tel phénomène faisant intervenir des variables déterministes et aléatoires et d'exploiter rationnellement ces données pour établir toutes relations de causalité par l'analyse et l'interprétation.

B. Portée et limitation de l'emploi de la statistique:

La statistique aura pour but d'étudier des faits pour permettre ensuite de prendre des décisions. On conçoit alors qu'elle puisse être dangereuse si l'étude été viciée par des erreurs d'observation; ou des erreurs d'interprétation.

Des critiques ont d'ailleurs été faites à la statistique:

On lui a reproché de ne porter que sur des faits passés, et en conséquence, d'apporter trop tard des enseignements, mais cela n'est pas vrai pour les domaines où justement apparaissent des permanences statistiques.

On a reproché aux statistiques d'être fausse; c'est vrai certaine mesure, mais il en est de toutes les bases numériques sur lesquelles ont été fondées les sciences expérimentales, la physique par exemple. Par ailleurs il est préférable d'avoir une connaissance même imparfaite des événements que pas des connaissances du tout.

On a dit de la statistique qu'elle aboutisse à des conclusions relatives au comportement des ensembles et non celui des individus, mais précisément le statisticien ne s'intéresse qu'à des ensembles et non à des unités.

On a accusé la statistique d'être une des formes de plus raffinées du mensonge. Un jugement sûr, et de la prudence, sont donc nécessaires en la matière avant de se livrer à toute conclusion. Il sera donc toujours indiqué d'établir soigneusement les statistiques qui seront utilisées, et de bien connaitre les techniques qui serviront à leur étude.

C- Histoire de la statistique:

Dans l'évolution de la statistique au cours de l'histoire on peut distinguer trois étapes : Dans cette étape qui va de la plus haute antiquité¬ jusqu'au 18 ième siècle, la statistique se réduit à des recensements et inventaires de caractère démographique ou comptable et reste une statistique de constatation.

- ✓ Ce n'est qu'aux 18 ième siècles que l'on voit apparaitre le rôle prévisionnel des statistiques et c'est à Adolphe Quételet que l'on doit l'idée que la statistique est une science s'appuyant sur les probabilités.
- ✓ Sous l'impulsion de grands mathématiciens, la troisième étape est révolutionnaire car elle va permettre désormais d'ouvrir à la statistique des horizons illimités, surtout avec l'essor de l'informatique qui va permettre de traiter et d'analyser un plus grand nombre de données.

D- Problème de Terminologie :

1- Différence entre la statistique et les statistiques :

a- Les statistiques :

Les statistiques représentent un ensemble de renseignement numériques découlant des recensements de population, des données de registre d´état- civil et enquêtes appropriées.

Les statistiques économiques sont par exemple, celle relatives aux prix, aux chiffres d´une entreprise, aux taux de croissance de l´économie, etc..

b- La statistique :

La statistique est un ensemble de techniques et de méthodes visant au rassemblement, á la présentation et á l´analyse des données quantitatives. Elle est donc une méthode scientifique donc l´objet est de recueillir, d´organiser, de résumer et d´analyser les données d´une enquête ou d´une étude. Elle permet également de tirer des conclusions logiques mais aussi de prendre des décisions á partir des analyses effectuées. Bien évidement lorsque nous voulons faire une étude statistique, l´idéal serait de prendre en considération tous les membres de la population. Seulement dans bien de cas, cela s´avère impossible en raison du manque d´argent ou de temps. Il convient donc dans ces situations de faire porter l´étude sur un ensemble plus restreint de la population.

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 données.

a) Population :

La population désigne un ensemble d'individus concernés lors de l'étude d'un sujet particulier. La population étudiée doit être définie avec précision.

Exemple1 : Si nous voulons étudier la population d'une ville, nous sommes tenus de préciser, si nous avons inclus dans cette population, les élèves internes des établissements scolaires, les malades de l'hôpital, les appelés du service militaire, les détenus en maison d'arrêt, etc.

Exemple 2 : Si nous voulons étudier le degré d'utilisation du téléphone, la population serait l'ensemble de toutes les personnes abonnées à celui-ci.

Exemple 3 :

- Etudier les notes de 1000 étudiants est une expérience statistique.
- Les unités statistiques ou bien les individus correspondent aux étudiants.
- La population est l'ensemble des étudiants.
- Lors de l'étude sur l'emploi la population d'étude pourrait être l'ensemble des personnes en âge de travaille.

Lorsque l'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 généraliser les résultats obtenus sur l'échantillon à toute la population.

b) individu :(unité statistique)

On appelle individu chaque élément de la population ou de l'échantillon (sous ensemble ou partie de la population). On utilise également le terme **unité statistique** pour désigner un individu.

Exemple 4 : Dans l'exemple précédent portant sur le degré d'utilisation du téléphone, chaque personne abonnée représente un individu ou une unité statistique.

c) Variable ou caractère :

Chaque individu d'une population est décrit par un ensemble de caractéristiques appelées variables ou caractères, on les représente souvent par des lettre majuscules : X, Y,….

Les valeurs qui peut prendre une variable statistique sont appelées modalités.

Une variable doit donc présenter au minimum deux modalités

Exemple 1:

Le caractère est l'aspect particulier que l'on désire étudier. Pour décrire les individus qui composent une population ou un échantillon, nous pouvons nous intéresser à leur âge, leur sexe, leur taille, leur religion, leur nationalité ou encore leur opinion sur un sujet donné.

Exemple 2 :

Du point de vue d'une entreprise, nous pouvons retenir le chiffre d'affaires, le bénéfice ou l'investissement de cette dernière.

d) Les modalités :

Les modalités sont les différentes situations dans lesquelles les individus peuvent se trouver à l'égard du caractère considéré.

Exemple 3 : Le sexe est un caractère qui présente deux modalités : féminin ou masculin.

Exemple 4 :

Si nous voulons étudier la satisfaction tirée par les étudiants de l'université de Béjaia de leurs dernières vacances estivales, les modalités peuvent être « peu », « modérément », « beaucoup » ou « énormément ».

Exemple 5 :

Si le caractère retenu ou la variable étudiée est le nombre d'enfants par familles, les modalités de ce caractère peuvent être 0,1, 2, 3, etc.

Exemple 6 :

Les modalités de la variable âge des ouvriers d'une entreprise peuvent être : [25,30[; [30,35[; [35,40[; [40, 45[; [45, 50[et [50, 60[.

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

1) **Variable quantitative :**

Une variable est dite quantitative si ses différentes valeurs sont mesurables. Il y a deux types de variables quantitatives :

a. Variable discrète :

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

Exemple 7 :

La variable nombre d'enfants par famille, dont les valeurs sont : 0, 1, 2, 3,4etc… est une variable quantitative discrète.

b. Variable continue :

Une variable quantitative est dite **continue** si elle prend toutes les valeurs dans des intervalles de l'ensemble des nombres réels $\mathbb{R}$.

Exemple 8 :

La variable taille en cm observée sur un échantillon de 500 individus est une variable quantitative continue, dont les résultats sont les suivants :

100 individus leurs tailles sont comprise entre dans l'intervalle [150, 160[; 250 individus dans l'intervalle [160, 170[; 150 dans l'intervalle [170, 185].

2) Variable qualitative :

La variable est dite qualitative si les différentes modalités ne sont pas mesurables.

Exemple 9:

Le sexe, la nationalité, l'état matrimonial, la couleur des yeux, la catégorie socioprofessionnelle.

<u>Cours 2 :</u>

e) La série statistique

On appelle série statistique, l'ensemble des différentes données associées à un certain nombre d'individus.

Exemple 1 : La série suivante provient d'une enquête auprès de quelques personnes pour connaître leur taille.

159 160 160 159 157 158 160 170 175
170 160 158 155 163 172 162 155

Exemple2 : La série suivante donne le salaire payé par une entreprise aux individus par rapport à leur sexe. L'enquête pourrait fournir la série statistique suivante :

Sexe	Féminin	Féminin	Masculin	Féminin	Masculin	Féminin
Salaires(DA)	15000	18000	30000	28000	22000	25000

Observations :

Une variable donne lieu à plusieurs observations. Si la variable est notée **X**, alors on notera par $x_1, x_2, x_3, \ldots x_n$ les observations ou série statistique, où **n** présente le nombre d'observations.

Si les observations portent sur la totalité de la population, on dit que **n** est la taille (ou l'effectif) de la population, 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 3 : Soit un échantillon de dix individus, la variable âge observé 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.

Données :

Lorsque les observations portent sur une ou plusieurs variables (on note p variable) et sur un ou plusieurs individus (on note n individus), l'ensemble des valeurs des observations récoltées constituent les données de l'étude ou de l'expérience effectuée. On peut les mettre sous forme d'un tableau individus/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_{n1}	………………	x_{2p}
. . .	. . .	. . .	. . .	. . .
I_n	X_{n1}	X_{n2}		X_{np}

Chaque individu **I** est représenté par un numéro d'ordre i=1, 2, ….n; chaque variable **X** est indicée d'un numéro d'ordre correspondant j=1, 2, ….p.

x_{ij} est la valeur prise par la**j^ème^**variable pour l'ième individu.

Exemple 4 :

Soit un échantillon de dix individus, sur lesquels on observe deux variables X_1 "le poids" et X_2 "l'âge", les résultats de cette étude sont représentés dans le tableau suivant.

Individus \ variables	X_1= poids	X_2= âge
I_1	40	12
I_2	50	18
I_3	65	25
I_4	75	30
I_5	80	42
I_6	72	37
I_7	78	25
I_8	68	38
I_9	65	19
I_{10}	64	24

Cours 3 :

Les tableaux statistiques

1. Présentation des données sous forme de tableaux :

Au cours d'une étude les données sont recueillies de façon désordonnée.

Cette masse d'informations que si les informations en question sont classées et mises en ordre, ce classement ne peut se faire que dans des tableaux statistiques qui servent de documentation statistique.

Ces tableaux doivent comporter:

- Le titre indiquant l'objet du tableau.
- L'unité utilisée (Dinars, milliers de tonnes, milliers de personnes, le pourcentage….).
- Les titres des lignes et des colonnes
- Les sources qui doivent être citées.

Leur présentation diffère selon la nature des variables

a. Cas d´un caractère qualitatif :

Pour présenter les résultats observés, nous distinguons d’abord les différentes modalités du caractère étudié (les différentes valeurs de la variable considérée). On les désigne par X1, X2, X3,… Xk. Il s’agit donc d’un caractère présentant k modalités.

Nous déterminons ensuite le nombre d’individus associés à chacune des modalités.

On parlera alors d’effectifs ou de fréquences absolues. Ils sont identifiés par n1, n2, n3,…,nk.

Nous pouvons aussi évaluer l’effectif relatif de chacune des modalités. On Parlera alors de fréquences relatives. Ces dernières sont représentées par *f,f,…,f k* 1

Nous pouvons également exprimer l’effectif relatif en pourcentage. Il suffit de multiplier l’effectif relatif de chacune des valeurs de la variable par **100.**

La présentation sous forme de tableaux consiste donc à confectionner un tableau contenant un titre et un corps susceptible de donner toutes les informations précédentes.

Individus n°	1	2	3	4	5	6	7	8	9	10
Nationalité	Etranger	Algérien	Etranger	Algérien	Algérien	Etranger	Algérien	Algérien	Etranger	Etranger

Pour présenter les résultats observés, nous distinguons d'abord les différentes modalités du caractère considéré. Le caractère étudié ici est la nationalité représentée par Xi. Nous déterminons ensuite le nombre d'individus associés à chacune des valeurs de cette variable. Ils sont identifiés par ni. Enfin, un tableau est construit pour résumer toutes ces informations.

Tableau: La répartition de 100 personnes selon leurs nationalité **.**

Nationalité *Xi*	**EffectifsAbsolus** n_i	**Effectifs Relatifs** *fi =n /N*	**EffectifsRelatifs** *fi(%)*
Algérien	6	0,6	60
Etranger	4	0,4	40
Ensemble	10	1	100

b- <u>Cas d´un caractère quantitatif :</u>

En raison de la nature des valeurs que peut prendre un caractère quantitatif dans une série statistique, nous distinguons deux (02) catégories : le caractère quantitatif discontinu (discret) et le caractère quantitatif continu

<u>Exemple 1</u> : Une enquête sur la destination de 12 touristes algériens a permet d'avoir les résultats suivants :

Touriste N°	Destination
1	Tunisie
2	Turquie
3	Turquie
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** a**lgériens**, elle présente **cinq modalités** : Tunisie, France, Turquie, Maroc, Espagne avec les **effectifs partiels** respectivement n_1=4, n_2=3, n_3=3, n_4=1, n_5=1.Delà, on obtient le tableau statistique correspondant :

Destination Xi	Effectifs partiels n_i	Fréquences f_i
Tunisie	4	0.333
France	3	0.25
Turquie	3	0.25
Maroc	1	0.083
Espagne	1	0.083
Total	12	1

Remarque :

Dans l'exemple précédent, la somme des fréquences n'est pas égale à 1, ceci est dû aux calculs effectués sur les fréquences. En effet on ne peut pas prendre tous les chiffres après la virgule, par contre on peut arrondir certain chiffre de manière à avoir la somme égale à 1.

a. Cas d'une variable discrète ou discontinue :

Lorsque la variable statistique est discrète, il arrive souvent qu'une même valeur x_i (i = 1, 2,3…..k; k =nombre de modalité) soit observée sur plusieurs individus, on note alors n_i son effectif (fréquence absolue), qui est le nombre d'observations égales à $\mathbf{x_i}$ et $\mathbf{f_i}$ la fréquence relative correspondante: $\mathbf{f_i} = {}^{n}i/n$, i= 1,2,…..k.

Le tableau statistique se présente alors sous la forme suivante:

x_i	n_i	$\mathbf{f_i}$

Remarque :

$$\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} \mathbf{f_i} = \mathbf{1}$$

Exemple 2: Distribution statistique du personnel d'une entreprise d'après le nombre d'enfants à charge.

Nombre d'enfants à charge	**Effectifs n_i**
0	4
1	15
2	29
3	18
4	10
5	3
6	1
Total	**80 = somme des n_i = $\sum_{i=1}^{i=7} n_i$ = N**

Dans cet exemple la variable discrète prend sept valeurs :

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

A chacune de ces valeurs correspond son effectif :

$n_1=4$ $n_2=15$ $n_3=29$ $n_4=18$ $n_5=10$ $n_6=3$ $n_7=1$

La série statistique (ou distribution statistique) est l'ensemble des sept couples :

(x_1, n_1) (x_2, n_2) (x_3, n_3) (x_4, n_4) (x_5, n_5) (x_6, n_6) (x_7, n_7)

Soit :

(0, 4) (1, 15) (2, 29) (3, 18) (4, 10) (5, 3) (6, 1)

Et nous dirons la série (x_i, n_i), l (indice i prenant les valeurs : 1, 2, 3, 4, 5, 6, 7.
La somme des effectifs est l'effectif total N de la population étudiée :

$$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7$$

$$\sum_{i=1}^{i=7} n_i$$

Que nous noterons:
Souvent nous nous contentons d'écrire $\sum \mathbf{n_i}$ (qu'on lira somme des n_i) pour noter la somme de tous les effectifs de la distribution statistique étudiée.

Exemple 3:

La distribution de fréquences suivante représente la répartition de 100 familles selon le nombre de d´enfants.

Tableau : La répartition de 100 familles selon le nombre d´enfants

Xi	*ni*	*Fi*
0	15	0,15
1	35	0,35
2	25	0,25
3	10	0,10
4	5	0,05
5	10	0,10
Ensemble	**100**	**1.00**

Exemple 4: La série résulte d´une enquête auprès de 30 étudiants pour connaitre leur âge

Tableau 2 : La répartition de 30 étudiants selon leur âge

Xi	*ni*	*fi*	*fi(%)*
17	3	0,1	10
18	8	**0,26**	**27**
19	6	0,20	20
20	4	0,13	13
21	4	0,13	13
22	3	0,1	10
23	2	**0,06**	**7**
Ensemble	**30**	**0,98 ≈ 1**	**100**

b-Cas d'une variable continue :

Lorsque la variable est continue, on regroupe les observations en **k** classes d'extrémités : c_0, c_1, c_2, c_3.....c_k

On note pour chaque classe [c_i, c_{i+1}[l'effectif n_i et la fréquence relative f_i, le tableau statistique se présente sous la forme suivante :

C_i C_{i+1}	n_i	f_i

$[c_i, c_{i+1}[$	n_i	f_i

Ou bien

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

Comme la variable statistique ne prend plus une valeur précise **Xi,** mais un ensemble de valeurs possible comprise dans un intervalle appelé classe, nous somme tenu de calculer le centre de classe, noté **ci** qui est la valeur de la variable statistique égale á la moyenne arithmétique des valeurs des extrémités de la classe.

On appelle centre d'une classe $[c_i, c_{i+1}[$ la valeur:

$$c_i = \frac{e_i + e_{i+1}}{2}$$

L´amplitude : d´une classe est la largeur de l´intervalle total de variation de la variable statistique, il peut également arriver que les classes extrêmes ne soient délimitées

D´où : et **l'amplitude** est notée $\mathbf{a_i} = e_{i+1} - e_i$

<u>Exemple 5</u> : On désir présenter la répartition de la population algérienne résidente de 2008

Tranche d'âge en années	Population en 2008
0-1	823000
1-4	1245700
4-9	1345200
9-13	1457000
. .	. .
75-80	225000
80 et plus	122500
total	23455000

Sources:ONS

Ainsi la lecture est facile et le regroupement peut se faire aussi selon une autre amplitude.

<u>Exemple 6</u> : 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.

Ces observations peuvent être représentées par des classes d'amplitude constantes égale à 4 cm par exemple :

Tailles	Effectif partiel n_i	Fréquences f_i	Fréquences en pourcentage
[158 – 162[	2	0.083	8.3%
[162 – 166[	6	0.25	25%
[166 – 170[	0	0	0%
[170 – 174[	4	0.167	16.7%
[174 – 178[	4	0.167	16.7%
[178 – 182[	6	0.25	25%
[182 – 186[	2	0.083	8.3%
Total	**24**	**1**	**100%**

Remarque :

Dans l'exemple précédent, la somme des fréquences n'est pas égale à 1, ceci est dû aux calculs effectués sur les fréquences. En effet on ne peut pas prendre tous les chiffres après la virgule, par contre on peut arrondir certain chiffre de manière à avoir la somme égale à 1.

Cours 4 :

Méthode de détermination de l´amplitude

Règle de STURGES :Cette règle est utilisée pour déterminer le nombre de classes à utiliser pour représenter une variable statistique continue.

Le nombre de classes est égal à l'entier le plus proche de la quantité :

$1 + 3.3 \log n_i$

Par exemple pour un échantillon de taille n = 200, on doit utiliser :

$1 + 3.3 \log 200 = 1+3.3\text{x}2.3 = 8.59 = 9$ classes

L'amplitude constante de ces classes sera égale à :

$$\boldsymbol{a} = \frac{\mathbf{X}\max - \mathbf{X}\min}{\textbf{nombre de classes}}$$

Exemple 1: Lors d´une journée, on relève les âges de 20 personnes venant se présenter á l´examen théorique du permis de conduite.

18- 19- 19- 23- 36- 21- 37- 23- 22- 19

18- 18- 20- 21- 19- 26- 32- 19- 21- 20

Règle de STURGES :

$$\boldsymbol{ai} = \frac{\mathbf{X}\max - \mathbf{X}\min}{\mathbf{1 + 3.3logN}}$$

D´ou :

$$\boldsymbol{ai} = \frac{\mathbf{37-18}}{\mathbf{1+3.3log20}} = \frac{\mathbf{19}}{\mathbf{5,31}} = \mathbf{4}$$

Donc on aura **5** classes d´amplitude **4**

Tableau :Répartition des candidats selon leurs âges

Xi	*ni*	*ai*	*fi*	*%*
[18-22[	13	4	0,65	65
[22-26[	03	4	0,15	15
[22-30[	01	4	0,05	05
[30-34[	01	4	0,05	05
[34-38]	02	4	0,10	10
∑	20	/	1	100

Cours 5 :

2- Présentation graphique des séries statistiques :

Soit une population d’un effectif total N que nous désirons étudier selon un certain caractère K. Nous désignons par ni l’effectif correspondant à chaque modalité du caractère. Comme nous l’avons vu précédemment, il est possible de présenter les résultats obtenus sous forme de tableaux. Mais il est également possible de recourir aux méthodes de représentation graphique. Celles-ci diffèrent selon le caractère qualitatif ou quantitatif de la variable étudiée.

1. **Cas d´une variable qualitative :**Une variable qualitative peut être représentée graphiquement de trois manières.

1.1. 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 veut calculer le degré du secteur représentant la modalité **i** d’effectif partiel ni alors on peut écrire :

n (l’effectif total) → 360°

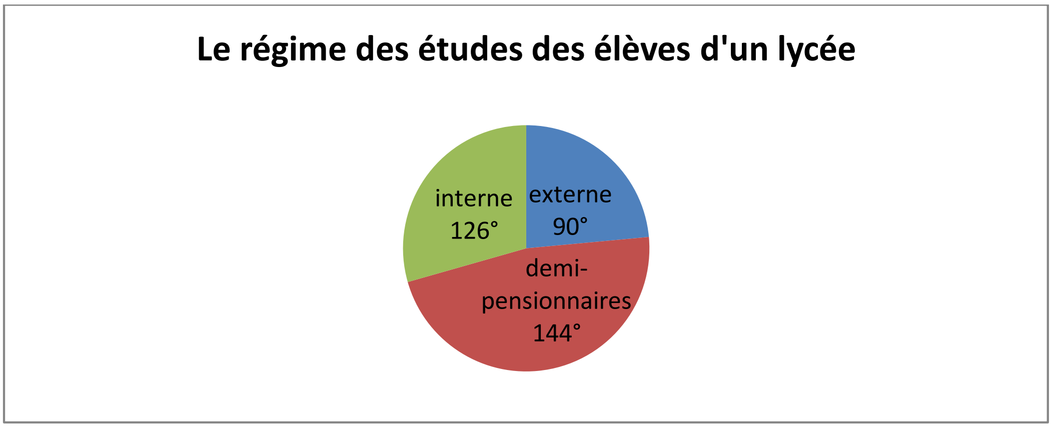
$$\rightarrow \mathrm{di} = \frac{\mathrm{ni \; x \; 360}}{\mathrm{n}} = f_i \text{ x } 360$$

Exemple 1 : On étudié le régime des études sur un échantillon de 200 élèves d’un lycée donné, les résultats obtenus sont les suivants :

Régime des études	Externe	Interne	Demi-pensionnaire
Nombred’élèvesn_i	70	50	80
Fréquences f_i	0.35	0.25	0.4

Calculons le degré correspondant aux différents secteurs :

- $d_i = f_i \times 360° = 0.35 \times 360° = 126°$
- $d_i = f_i \times 360° = 0.25 \times 360° = 90°$
- $d_i = f_i \times 360° = 0.4 \times 360° = 144°$



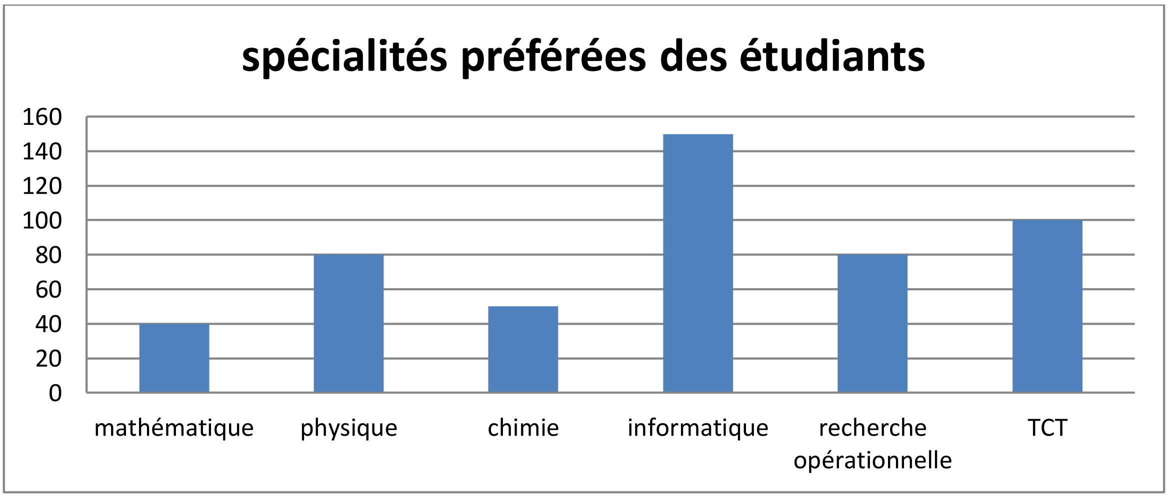
2.2 Représentation en tuyau d'orgue :

Elle consiste à porter en abscisse les modalités de la variable étudiée, au-dessus des quelles on trace des tuyaux d'orgues de même largeur, et en ordonnée les effectifs ou les fréquences qui vont déterminer leurs auteurs.

Exemple 2: Une enquête auprès de 500 étudiants de 1ère année SETI (sciences exactes, technologie et informatique) admis en 2ème année sur les spécialités préférées a fourni les résultats suivants :

Spécialité préférée	Nombre d'étudiants
mathématique	40
Physique	80
Chimie	50
informatique	150
rechercheopérationnelle	80
TCT	100
Total	500

Ces résultats peuvent être traduits par un diagramme de tuyaux d'orgue suivant

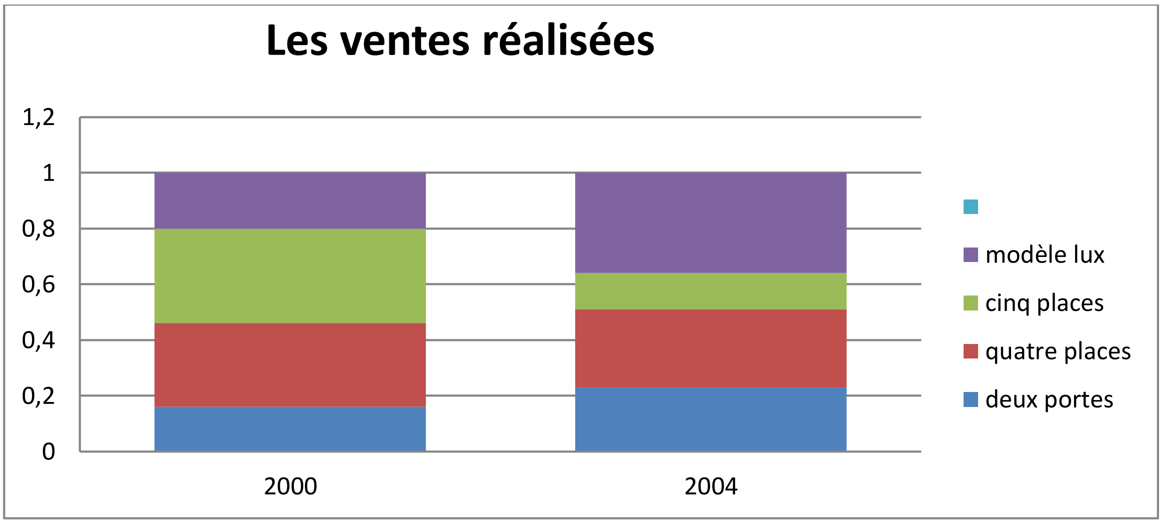


3.3 Diagramme en bandes :

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

Exemple 3: Les ventes réalisées par une maison de fabrication d'automobiles durant les années 2000 et 2004 sont les suivantes :

Véhicules	2014		2018	
	n_i	f_i	n_i	f_i
Deuxportes	800	0.16	1600	0.23
Petites 4 places	1500	0.3	2000	0.28
Cinq places	1700	0.34	900	0.13
Modèleluxe	1000	0.2	2500	0.36
Total	5000	1	7000	1



2. Cas d´une variable quantitative :

On va distinguer les diagrammes correspondant à une variable discrète et ceux correspondant à une variable continue.

2.1. Le diagramme en bâton :

Il consiste à représenter les divers valeurs $\mathbf{x_i}$ prise par une variable discrète, en fonction des fréquences $\mathbf{f_i}$ ou les effectifs $\mathbf{n_i}$ de la manière suivante :

On représente en abscisse les observations $\mathbf{x_i}$ et au-dessus de chaque valeur en trace un bâton dont la hauteur et proportionnelle à $\mathbf{f_i}$ ou $\mathbf{n_i}$.

Exemple 4: Les poids (en kg) de 200 individus sont représentés dans le tableau suivant :

x_i	55	60	65	70	75	80	Total
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 bâton en utilisant les effectifs n_i ou les fréquences f_i.

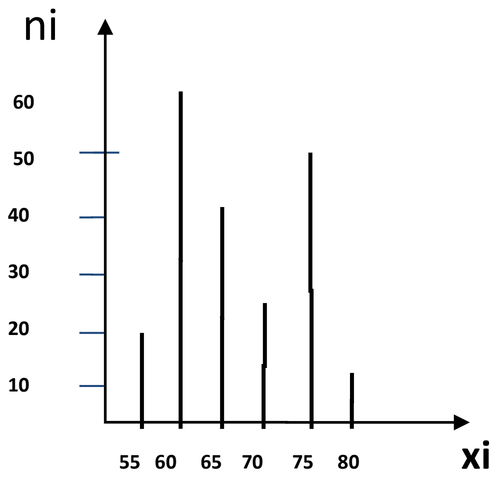
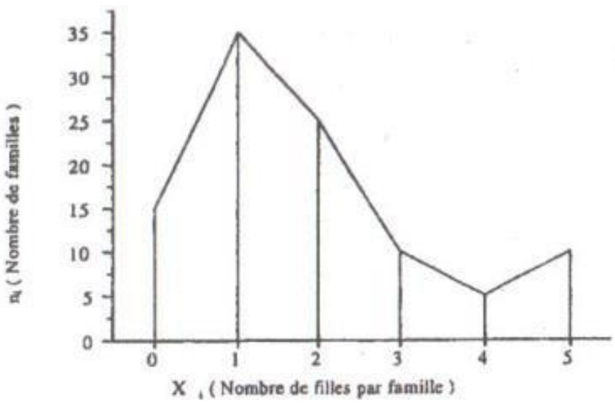


Diagramme en bâton

A chaque valeur de la variable Xi portée en abscisse, on fait correspondre un bâton de longueur proportionnelle à l'effectif ni correspondant à cette valeur.

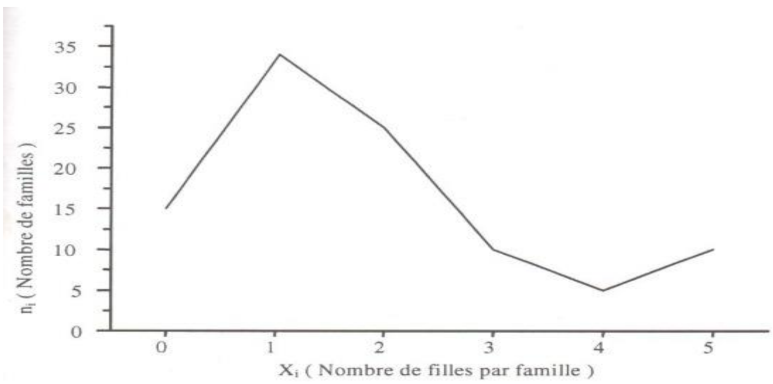
Polygone d'effectifs :

Il suffit de relier par un segment de droite les sommets de bâtons obtenus antérieurement



Courbe de fréquences

Il s'agit ici tout simplement d'un polygone de fréquences dont les différents sommets arrondis



Courbe cumulative croissante et décroissante

Dans ce cas, nous traçons une courbe dont seuls quelques ponts sont connus.

*** Pour la courbe croissante,** on joint les points de coordonnées (*i*ei, *n)*

c'est-à-dire :

- ✓ en abscisse, la valeur de l'extrémité droite ei de la classe i
- ✓ -en ordonnée, la valeur de l'effectif cumulé ↑ *i n* correspond cette classe.

*** Pour la courbe décroissante,** on joint les points de coordonnées *(e*,*n* 1)

c'est-à-dire :

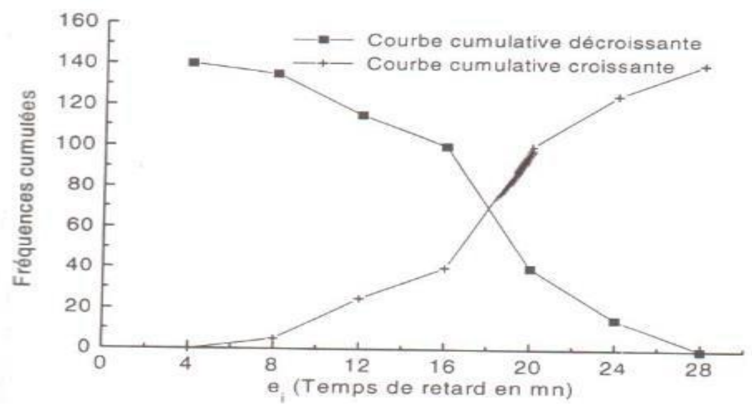
- ✓ en abscisse, la valeur de l'extrémité gauche e_{i-1} la classe i ;
- ✓ en ordonnée, la valeur de l'effectif cumulé ni ↓correspond cette classe.

Exemple 5: Cas d'une variable dont les classes sont égales.

Tableau : La répartition des 140 ouvriers selon leur temps de retard (le retard est exprimé en minutes)

Xi	**ni**	ai	ci	n_i↑ (E cc)	n_i↓(E cd)
[4-8[	5	4	6	5	140
[8-12[	20	4	10	25	135
[12-16[	15	4	14	40	115
[16-20[	60	4	18	100	100
[20-24[	25	4	22	125	40
[24-28[	15	4	26	140	15
Total	140	/	/	/	/

La courbe cumulative croissante et décroissante



Remarque : la même chose pour le cas d'une variable dont les classes sont inégales on trace les courbes de fréquences cumulées sans corriger l'effectif.

Attention : c'est une courbe que l'on trace et non une succession de segments de droite. Par ailleurs, on remarquera que les courbes croissantes et décroissantes sont symétriques par rapport à un axe parallèle à l'axe des abscisses et ayant comme ordonnée N/2

2.2 Histogramme :

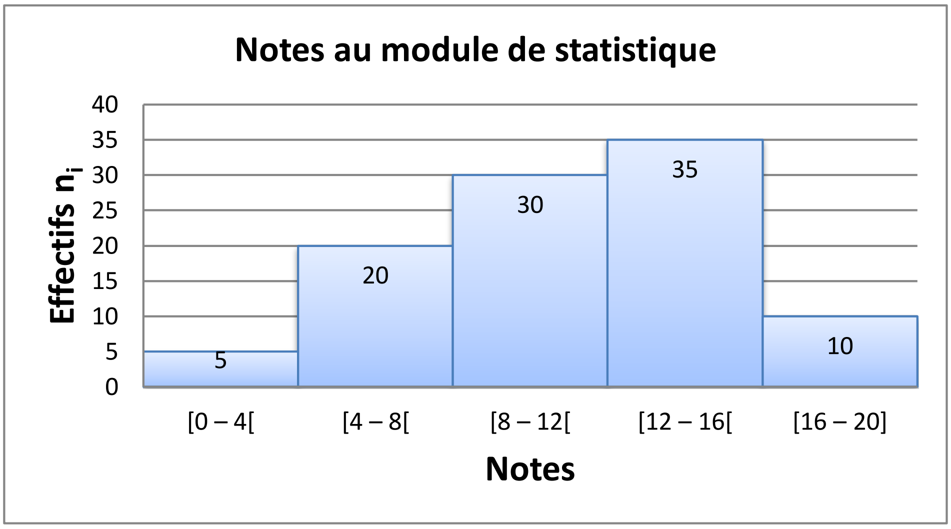
L'histogramme est utilisé pour représenter une variable continue.

On représente sur l'axe des abscisses les différentes classes supposée de même amplitude et au-dessus de chacune d'elle on trace des rectangles, dont les hauteurs sont discrètement proportionnelles aux fréquences ou aux effectifs.

Exemple 6: Les notes au module de statistique observées sur un échantillon de 100 étudiants sont données dans le tableau suivant :

Notes Xi	[0 – 4[	[4 – 8[	[8 – 12[	[12 – 16[	[16 – 20]	Total
Effectifsn_i	5	20	30	35	10	100
ai	4	4	4	4	4	/
cl	2	6	10	14	18	/
Ecc	5	25	55	90	100	/
Ecd	100	95	75	45	10	/

La variable étudiée est continue, donc elle peut être représentée graphiquement au moyen de l'histogramme suivant :



Remarque:

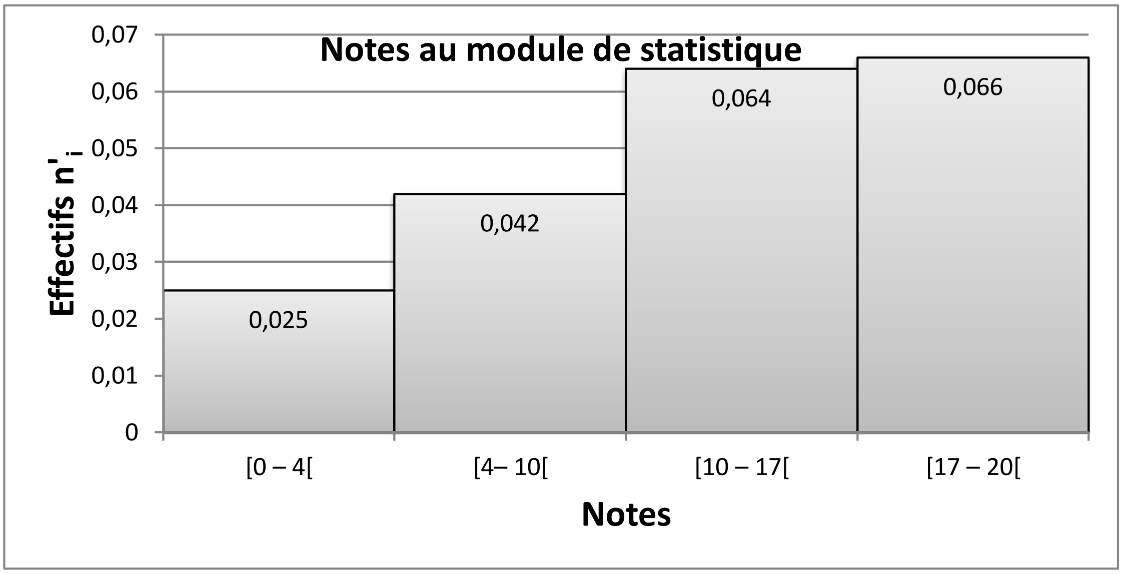
Ce mode de représentation est valable seulement dans le cas où on a des classes d'intervalles d'amplitude constante.

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 noté ni° :

$$\mathbf{ni°} = \frac{\boldsymbol{lapluspetiteamplitude}}{\boldsymbol{amplitudedelaclassei}}$$

Exemple 7 :

Notes	[0 – 4[	[4 – 10[	[10 – 17[	[17 – 20]	Total
Effectifsn_i	10	25	45	20	100
Amplitude$\boldsymbol{a_i}$	4	6	7	3	/
Effectifsrectifié$\boldsymbol{n_i}$	$10 \times \frac{\mathbf{3}}{\mathbf{4}} = \mathbf{7}$	$25 \times \frac{\mathbf{3}}{\mathbf{6}} = \mathbf{13}$	$45 \times \frac{\mathbf{3}}{\mathbf{7}} = \mathbf{19}$	$20 \times \frac{\mathbf{3}}{\mathbf{3}} = \mathbf{20}$	/



On peut aussi représenter l'histogramme autrement, en utilisant comme hauteurs des rectangles les quantités : $\frac{\mathbf{fi}}{\mathbf{ai}}$**ou (**$\frac{\mathbf{ni}}{\mathbf{ai}}$**).**

Notes	[0 – 4[	[4 – 10[	[10 – 17[	[17 – 20]	Total
Effectifsn_i	10	25	45	20	100
Fréquences f_i	0.1	0.25	0.45	0.2	1
Amplitude$\boldsymbol{a_i}$	4	6	7	3	/
Fréquencesrectifié$\frac{\mathbf{fi}}{\mathbf{ai}}$	$\frac{0.1}{4} = 0.025$	$\frac{\mathbf{0.25}}{\mathbf{6}} = \mathbf{0.042}$	$\frac{\mathbf{3}}{\mathbf{7}} = \mathbf{0.064}$	$\frac{\mathbf{0.2}}{\mathbf{3}} = \mathbf{0.066}$	/