

---

# Statistique Descriptive

Pr HAMDİ FEKKAK

---

# Objectifs

- ***Connaître les grands aspects de la statistique ainsi que son utilité***
- ***Apprendre les principales techniques de statistique descriptive univariée et bivariée.***
- ***Être capable de mettre en œuvre ces techniques de manière appropriée dans un contexte donné.***

# Références

- Dodge Y.(2003), Premiers pas en statistique, Springer.
- Dreesbeke J.-J. (1997), Eléments de statistique, Editions de l'Université libre de Bruxelles/Ellipses.
- Étienne B. Jean-Claude K. (2009), Statistique descriptive avec Excel et la calculatrice, Université Paris 1 Panthéon-Sorbonne

# Statistique

***« C '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. »***

# QU'EST CE QUE LA STATISTIQUE DESCRIPTIVE?

*La Statistique Descriptive est l'ensemble des méthodes et techniques permettant de présenter, de décrire et de résumer des données nombreuses et variées.*

# PLAN

# Partie I : Statistique descriptive cas d'une série univariée

## Chapitre 1 : définitions , typologie et représentations graphiques

- Définitions et vocabulaire
- Typologie des variables et représentations graphiques
  - Les variables qualitatives
  - Les variables quantitatives

## Chapitre 2: Paramètres statistiques

- Les paramètres de position
- Les paramètres de dispersion

# Partie II: Statistique descriptive cas d'une série bivariée

## Chapitre 3: distribution à deux caractères

- Distributions marginales
- Distributions conditionnelles
- Indépendance statistique
- Caractéristiques marginales et conditionnelles

## Chapitre 4: Analyse bidimensionnelle

- Covariance
- Coefficient de corrélation linéaire
- Régression linéaire



# Chapitre 1

## Définitions, typologie et représentations graphiques

# I- Définitions et vocabulaires

## 1-1- Statistique descriptive

- **La statistique descriptive** est l'ensemble des méthodes et techniques mathématiques permettant de présenter, de décrire et de résumer des données nombreuses et variées.
- La statistique descriptive s'applique à la plupart des disciplines :Agronomie, Biologie, Démographie, Economie, Sociologie,...

# I- Définitions et Vocabulaire

Exemple:

- Considérons les 100 familles d'une région donnée ;
- On tire au hasard 20 familles parmi ces 100 familles ;
- on interroge ces 20 familles tirées pour connaître leur nombre d'enfants.

Famille	Nombre d'enfants	Famille	Nombre d'enfants
F1	1	F11	0
F2	2	F12	1
F3	3	F13	2
F4	2	F14	1
F5	0	F15	2
F6	2	F16	2
F7	4	F17	0
F8	1	F18	2
F9	1	F19	2
F10	2	F20	3

Les 100 familles de la région :

Population statistique

Les 20 familles tirées au hasard :

Echantillon

Une Famille:

individu

Nombre d'enfant:

Caractère

0  
1  
2  
3  
4

Modalités

# I- Définitions

## 1-2- Population statistique :

- La population statistique est l'ensemble des personnes, des objets, des éléments sur lesquels on veut effectuer l'étude.
- Exemples :
  - ensemble de personnes interrogées pour une enquête
  - ensemble de surfaces cultivées sur lesquelles on mesure un rendement
  - ensemble de pays pour lesquels on dispose de données démographiques, géographiques ou économiques, ...

# I- Définitions

## 1-3- Un échantillon:

Un échantillon est un ensemble d'éléments tirés de la population , au hasard , sur lequel on effectue une étude exhaustive pour ensuite porter certaines conclusions sur l'ensemble de la population .

### Exemple

30 étudiants tirés au hasard parmi les étudiants de la section 5 Sc. Éco. Niveau 1ère année.

# I- Définitions

## 1-4- Individu (ou unités statistiques) :

Les individus sont les éléments de la population statistique étudiée.

Pour chaque individu, on dispose d'une ou plusieurs observations.

### Exemples :

- chacune des personnes interrogées pour une enquête
- chaque parcelle cultivée en vue d'étudier le rendement
- chaque pays pour lequel on étudie des données socio-économiques, ..

# I- Définitions

## 1-5- Caractère statistique (ou Variable statistique)

C'est ce qui est observé ou mesuré sur les individus d'une population statistique.

Exemples :taille, poids, salaire, rendement, température moyenne, sexe, CSP,...

# I- Définitions

## 1-6- Modalités :

Les modalités d'une variable sont les différentes valeurs que peut prendre celle-ci.

### Exemple

-les modalités de la variable "situation familiale" sont : célibataire, marié, veuf, divorcé.

-Les modalités de la variable "sexe" sont : féminin, masculin.



# I- Définitions

## 1-7- Effectif et fréquence

On appelle *effectif d'une* modalité , le nombre de fois que cette modalité apparait.

On définit la fréquence d'une modalité comme le rapport de son effectif à l'effectif total .

Modalités	Effectif	Pourcentage (F)
0	3	15% (3/20)
1	5	25% ( 5/20)
2	9	45% (9/20)
3	2	10% ( 2/20)
4	1	5% (1/20)
Total	20	100% (1)

Un tableau  
d'effectifs

# Application1

Afin de connaître le nombre actuel d'étudiants par classe pour l'ensemble des classes d'une certaine école, on note ce nombre pour 30 d'entre elles choisies au hasard .

L'ensemble de ces observations se présente comme suit :

30	27	30	26	27	33	26	28	28	31
32	30	26	30	32	31	27	33	34	27
32	28	31	27	28	32	26	27	30	32

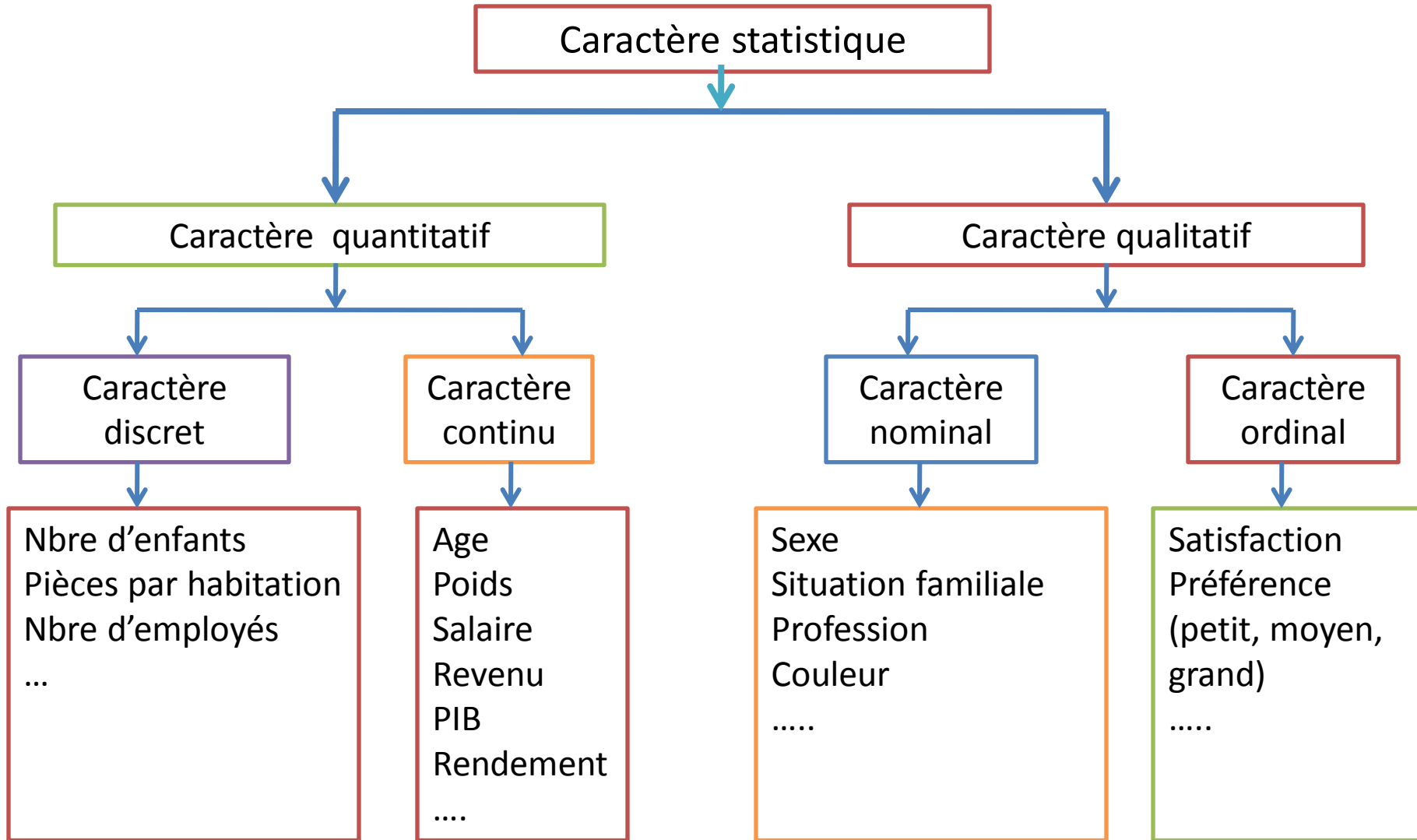
– Identifier : la population ,l'échantillon, l'individu ,le caractère étudié , les modalités ,les effectifs et les fréquences (pourcentages).

## Réponse

- La population : l'ensemble des classes de l'école.
- Un individu : chacune de ces classes
- L'échantillon observé: 30 classes choisies au hasard
- Le caractère étudié : le nombre actuel d'étudiants par classe

Modalités	26	27	28	30	31	32	33	34	Total
Effectifs	4	6	4	5	3	5	2	1	30
Pourcentage	13,3%	20%	13,3%	16,7%	10%	16,7%	6,7%	3,3%	100%

## II- Typologie du caractère( variable) statistique



## II- Typologie de Caractères statistiques (des variables)

### 2-1-Caractère qualitatifs

**Un caractère statistique est dit qualitatif lorsque son observation ne peut pas être traduite par une mesure; ses diverses modalités sont simplement constatées , repérées par un mot traduisant un « état ».**

- **Le caractère qualitatif peut être :**
  - **Nominal (nom)**
  - **ou Ordinal .(suit un ordre )**

# II- Typologie de variables

## 2-1- caractères qualitatifs (variables catégoriques)

### 2-1-1- Un caractère nominal

- **Un caractère nominal** décrit un nom ou une catégorie . Les noms ou les catégories possibles ne suivent pas un ordre naturel.
- Exemples:
  - L'état civil.
  - Genre
  - Profession
  - Filière
  - ...

- **A- Tableau brut d'un caractère nominal**

Exemple: Considérons la série statistique relative à 20 employés d'une entreprise donnée :

Employés	1	2	3	4	5	6	7	8	9	10
Etat Civil	M	M	D	C	C	M	C	M	M	C
Employés	11	12	13	14	15	16	17	18	19	20
Etat Civil	C	M	V	M	V	M	C	C	M	D

### Tableau d'effectifs

Le tableau d'effectifs est un tableau qui nous permet de synthétiser les modalités d'un caractère statistique, les effectifs et les fréquences de celles-ci.

Les modalités $X_i$	$n_i$	$f_i$
M	9	0,45
C	7	0,35
D	2	0,1
V	2	0,1
Total	$\sum n_i = 20$	$\sum f_i = 1$

## B- Représentation graphique d'un caractère nominal

Le tableau statistique d'un caractère qualitatif nominal peut être représenté par deux types de graphique:

- Les effectifs sont représentés par un diagramme en barres
- et les fréquences par un diagramme en secteurs

Diagramme en secteurs des fréquences

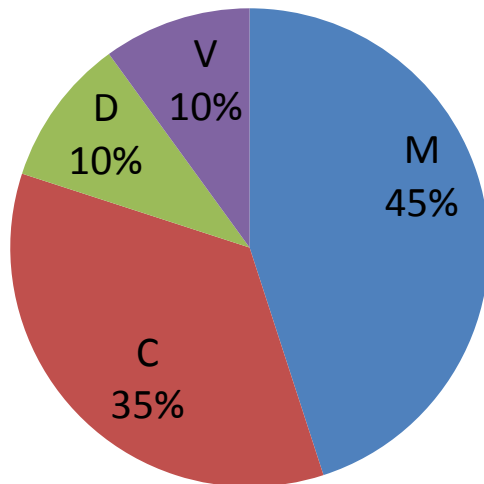
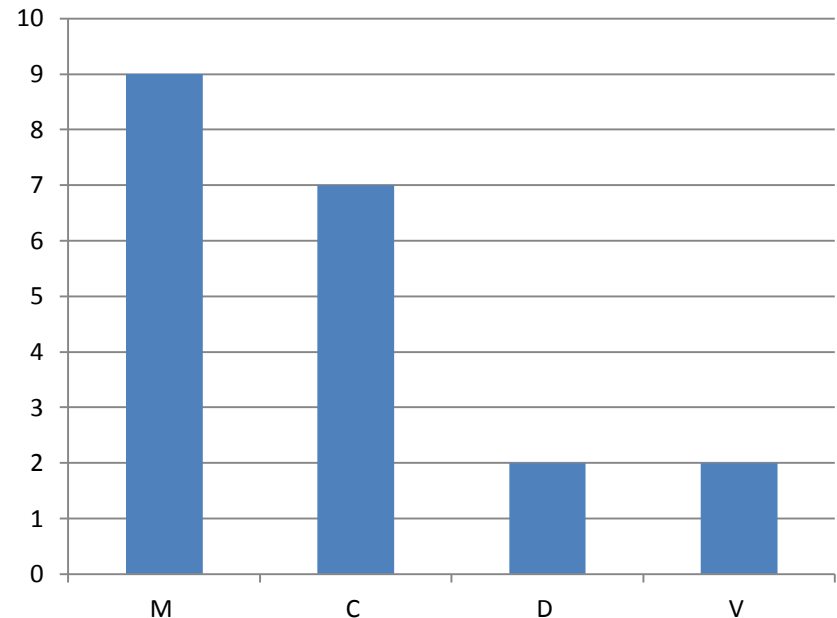


Diagramme en barres





# II- Typologie de variables

- **2-1- caractère qualitatifs**
- **2-1-2- Un caractère ordinal** est un caractère dans lequel les catégories possibles peuvent être classées dans un ordre spécifique ou dans un ordre naturel quelconque
- Exemple:
  - Préférences : Excellent , très bon , bon, moyen , faible
  - Mentions : Passable, Assez bien ,Bien , Très bien

# A- Le tableau statistique

**Exemple :**

**On interroge 30 employés d'une entreprise ,tirés au hasard, sur leur dernier diplôme obtenu.**

*Soit la série statistique de la variable « dernier diplôme »*

S.Dip	Sup	Prim	Qual	Sec	Prim	Prim	Sec	S.dip	Sup
Qual	Sec	Prim	Qual	Sec	S.dip	S.dip	Sup	Prim	Prim
Qual	Sec	Prim	Qual	S.dip	Sec	Prim	Sec	Qual	Sup

Modalités $X_i$	Effectifs ( $n_i$ )	Fréquences ( $f_i$ )
S.dip	5	0,17
Prim	8	0,27
Sec	7	0,23
Qual	6	0,2
Sup	4	0,13
Tot	30	1

# B- Représentations graphiques

Diagramme en barres

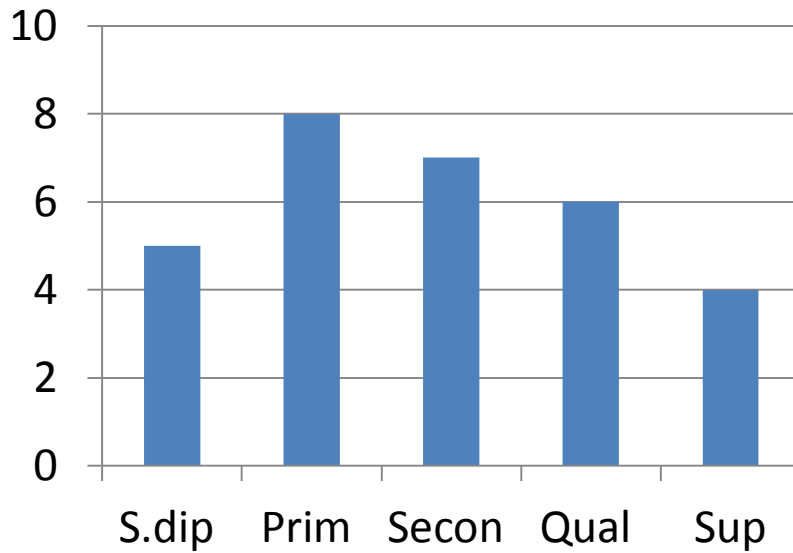
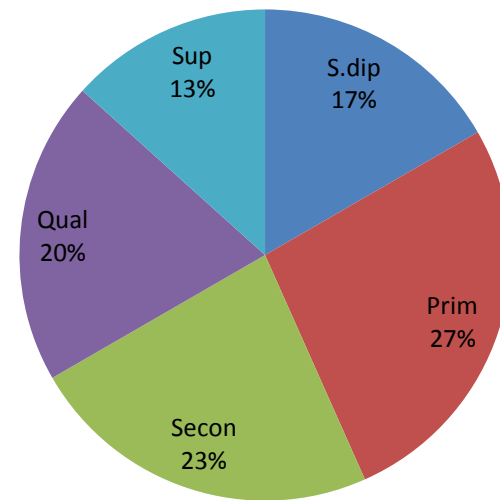


Diagramme en secteur des fréquences



# II- Typologie de variables

## 2-2- Variables quantitatives

- **Une variable quantitative**, connue aussi sous le nom de variable **numérique**, est une variable qui peut supposer un nombre infini ou défini de valeurs réelles.
- Les variables **quantitative** peuvent être **continues** ou **discrètes**.

# II- Typologie de variables

## 2-2- Variables quantitatives

### 2-2-1- Variables discrètes

- Une variable discrète a un domaine dénombrable c. à.d l'ensemble de ses valeurs possibles est dénombrable

Exemple :

Membres d'un ménage.

Nombre d'enfant par famille

Nombre de pièces par habitation

## Considérons

- $X$  : le caractère étudié,
- $n$  : le nombre d'observations pour ce caractère,
- $x_1, \dots, x_k$  : l'ensemble des modalités prises par ce caractère. On les écrit en ordre croissant si elles sont ordonnables.

Pour chaque modalité  $x_i$ , on note

- $n_i$  = effectif (ou fréquence absolue) de  $x_i$   
= nombre d'individus pour lesquels le caractère  $X$  prend la valeur  $x_i$ .
- $f_i$  = fréquence (relative) de  $x_i = \frac{n_i}{n}$   
= proportion d'individus pour lesquels le caractère  $X$  prend la valeur  $x_i$ .
- $N_i$  = effectif cumulé croissant de  $x_i$   
=  $n_1 + n_2 + \dots + n_i$   
= nombre d'individus pour lesquels le caractère  $X$  prend une valeur  $\leq x_i$ .
- $N'_i$  = effectif cumulé décroissant de  $x_i$   
=  $n_i + n_{i+1} + \dots + n_k$   
= nombre d'individus pour lesquels le caractère  $X$  prend une valeur  $\geq x_i$ .



## Remarques

i)  $n_1 + \dots + n_k = n.$

ii)  $0 \leq f_i \leq 1$  et  $f_1 + \dots + f_k = 1.$

iii)  $p_i = f_i \times 100 =$  pourcentage (ou fréquence relative en pourcentage).

iv) On définit de la même façon

- $F_i = f_1 + f_2 + \dots + f_i = \frac{N_i}{n}.$

- $P_i = p_1 + p_2 + \dots + p_i = F_i \times 100.$

v) Les  $N_i$ ,  $F_i$ ,  $N'_i$  et  $F'_i$  n'ont de sens que si on peut ordonner les modalités.

$X$	$n_i$	$f_i$	$N_i$	$F_i$
$x_1$	$n_1$	$f_1$	$N_1 = n_1$	$F_1 = f_1$
$x_2$	$n_2$	$f_2$	$N_2 = n_1 + n_2$	$F_2 = f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$n_i$	$f_i$	$N_i = n_1 + \dots + n_i$	$F_i = f_1 + \dots + f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k$	$N_k = n_1 + \dots + n_k$	$F_k = f_1 + \dots + f_k$
<b>Total</b>	$n$	<b>1</b>	—	—

## Remarques

- i) Pour une variable quantitative continue (ou traitée comme telle),  $x_i$  représente le centre de la  $i^{\text{ème}}$  classe.
- ii) On peut aussi inclure dans le tableau précédent les deux colonnes donnant les effectifs et les fréquences cumulé(e)s décroissant(e)s  $N'_i$  et  $F'_i$ .





$X$	$n_i$	$f_i$	$N'_i$	$F'_i$
$x_1$	$n_1$	$f_1$	$N'_1 = n_1 + \dots + n_k$	$F'_1 = f_1 + \dots + f_k$
$x_2$	$n_2$	$f_2$	$N'_2 = n_2 + \dots + n_k$	$F'_2 = f_2 + \dots + f_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$n_i$	$f_i$	$N'_i = n_i + \dots + n_k$	$F'_i = f_i + \dots + f_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$n_k$	$f_k$	$N'_k = n_k$	$F'_k = f_k$
<b>Total</b>	$n$	<b>1</b>	—	—

# Exemple « Nombre d'enfants par famille »

Famille	Nombre d'enfants	Famille	Nombre d'enfants
F1	1	F11	0
F2	2	F12	1
F3	3	F13	2
F4	2	F14	1
F5	0	F15	2
F6	2	F16	2
F7	4	F17	0
F8	1	F18	2
F9	1	F19	2
F10	2	F20	3

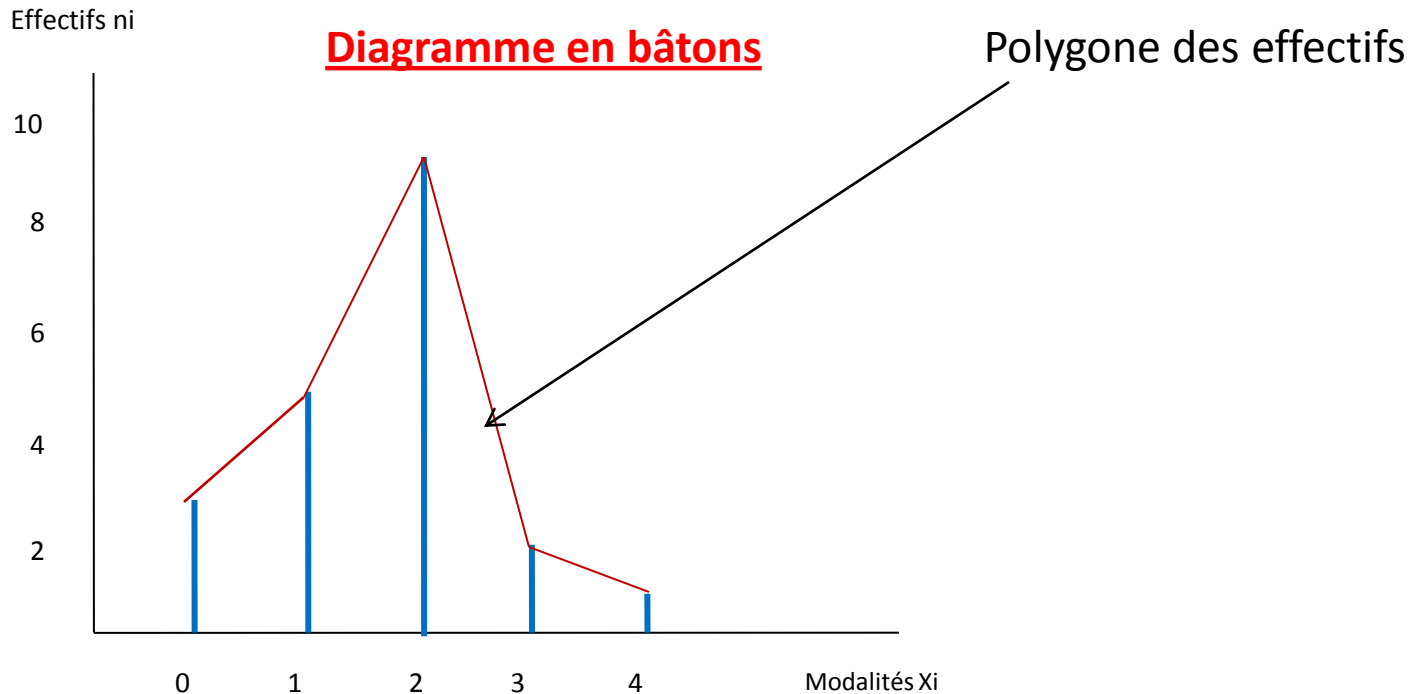
# A- Le tableau statistique

- On peut calculer les effectifs, les effectifs cumulés, les fréquences, les fréquences cumulées.

Modalités $X_i$	Effectifs $n_i$	Effectifs Cumulés croissant ( $N_i$  )	Effectifs Cumulés décroissant ( $N_i$  )	Fréquences $f_i$	Fréquences Cumulées croissant ( $F_i$  )	Fréquences Cumulées décroissant ( $F_i$  )
0	3	3	20	0,15	0,15	1
1	5	8	17	0,25	0,40	0,85
2	9	17	12	0,45	0,85	0,60
3	2	19	3	0,1	0,95	0,15
4	1	20	1	0,05	1	0,05
Total	20	---	---	1	---	---

# B- Représentations graphiques

- Quand la variable est discrète, les effectifs sont représentés par des bâtonnets





# C- Représentations graphiques des fréquences ( effectifs) cumulées croissantes

Les fréquences cumulées sont représentées au moyen de la fonction de répartition  $F(x)$ .

Cette fonction est définie de  $R$  dans  $[0, 1]$  comme suit :

Soit  $x_1, x_2, \dots, x_k$  les modalités d'une variable statistique donnée et  $f_i$  la fréquence relative à la modalité  $x_i$ .

Alors :

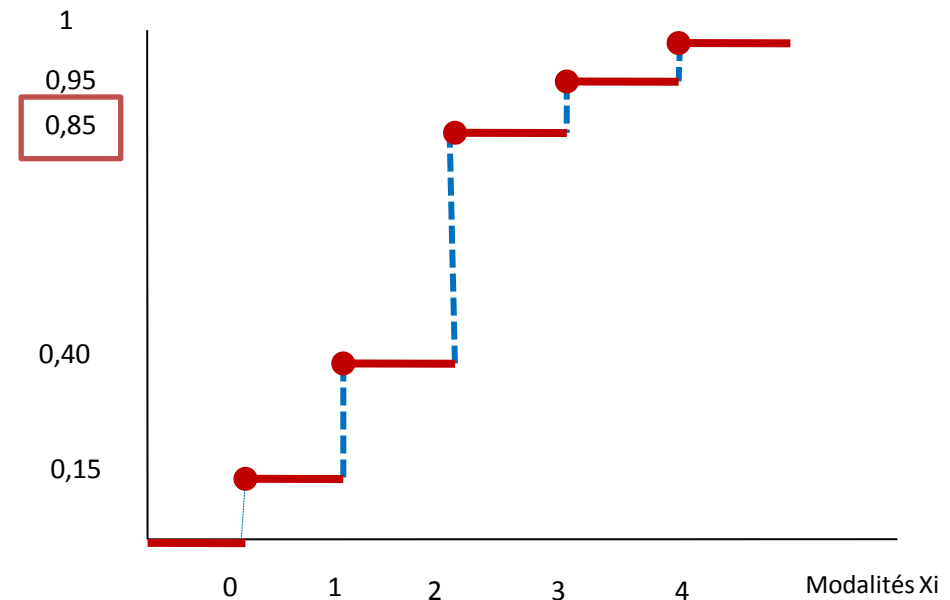
$$F(x) = 0 \quad \text{si } x < x_1$$

$$F(x) = \sum_{i=1}^k f_i \quad \text{si } x_1 \leq x \leq x_k$$

$$F(x) = 1 \quad \text{si } x \geq x_k$$

**Avec  $F$**  est constante sur chaque intervalle séparant deux modalités consécutives

Fréquences cumulées croissantes  $F_i$



On dit que 85 % des familles de la région ayant au plus 2 enfants par famille.

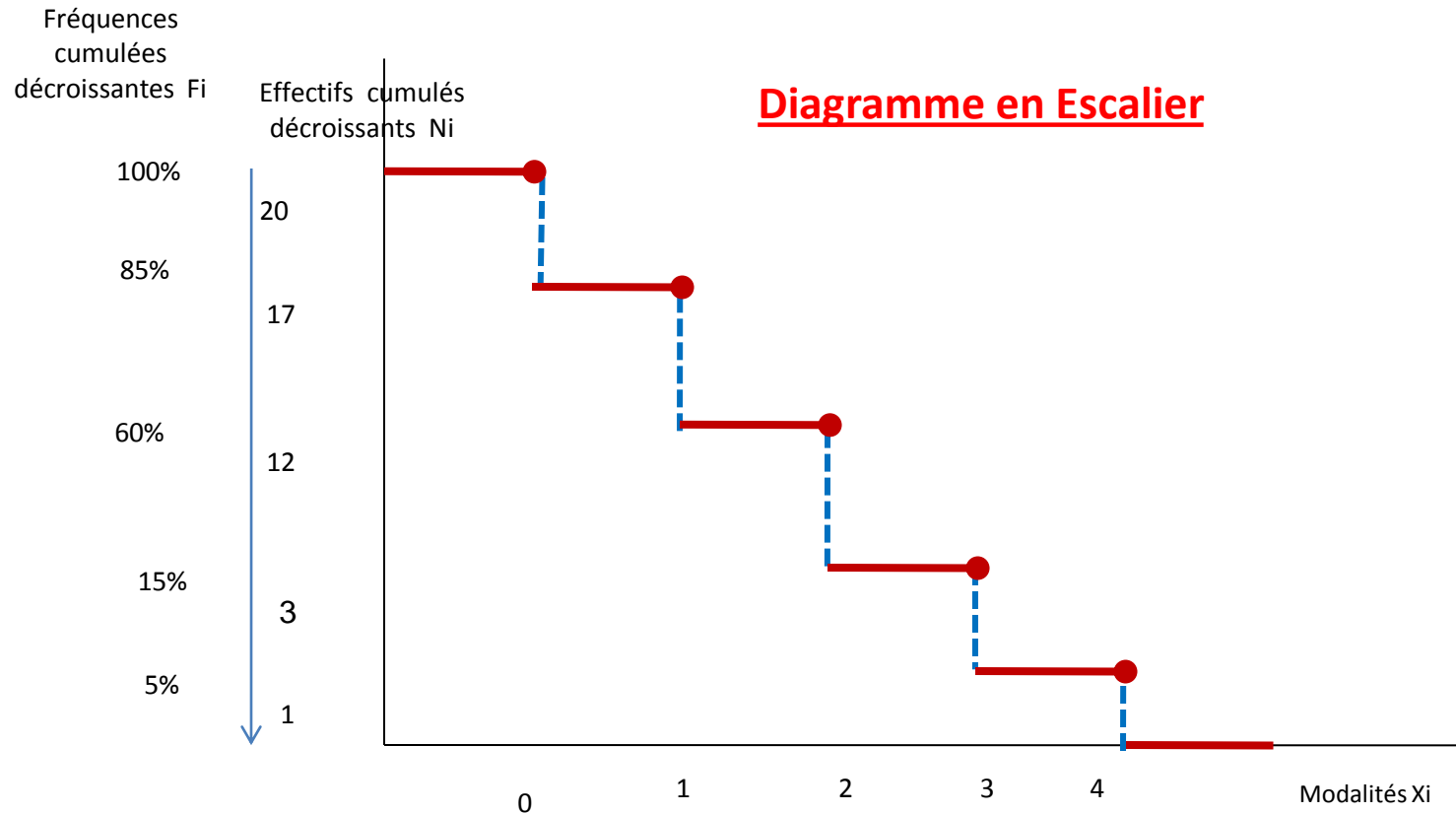


## Remarque

On utilise le polygone des effectifs (ou fréquences) cumulé(e)s croissant(e)s pour répondre graphiquement aux questions du style :

Quel est le nombre (ou pourcentage) d'individus dont la valeur du caractère est inférieure ou égale à  $x$  ?

# D- Représentations graphiques des fréquences (effectifs) cumulées décroissantes



On dit que 60 % des familles de la région ayant au moins 2 enfants par famille.

## Remarque

On utilise le polygone des effectifs (ou fréquences) cumulé(e)s décroissant(e)s pour répondre graphiquement aux questions du style :

Quel est le nombre (ou pourcentage) d'individus dont la valeur du caractère est supérieure ou égale à  $x$  ?

## 2-2- Variables quantitatives

### 2-2-1- Variables continues

- Une variable continue peut prendre toutes les valeurs d'un intervalle.
- Pour étudier une variable statistique continue, on définit des classes ou intervalles de valeurs possibles.

## 2-2- Variables quantitatives

### 2-2-1- Variables continues

• On appelle **amplitude de la classe**  $[a_i ; b_i[$  le réel noté  **$A_i$**  représentant la longueur de l'intervalle et définie par :  $A_i = b_i - a_i$ .

Avec  $a_i$  et  $b_i$  sont respectivement les bornes inférieure et supérieure de la classe  $n_i$ .

• Le **centre de la classe**  $[a_i ; b_i[$  est le réel noté  **$C_i$**  représentant le milieu de l'intervalle et donné par :  $C_i = (a_i + b_i) / 2$ .

# Exemple

On mesure la taille en centimètres de 50 élèves choisis au hasard d'une section donnée :

152	154	156	157	159	161	162	164	168	170
152	154	156	157	159	160	162	164	168	171
152	154	156	157	160	160	163	165	168	171
153	155	156	158	160	161	164	166	169	171
153	155	156	158	160	162	164	167	169	171

- Présenter le tableau statistique en prenant des intervalles égaux de 5 cm
- Représenter graphiquement cette distribution statistique

# A-tableau statistique

## \* 1<sup>er</sup> Cas : les amplitudes sont égales

- Soit les modalités (classes) suivantes:

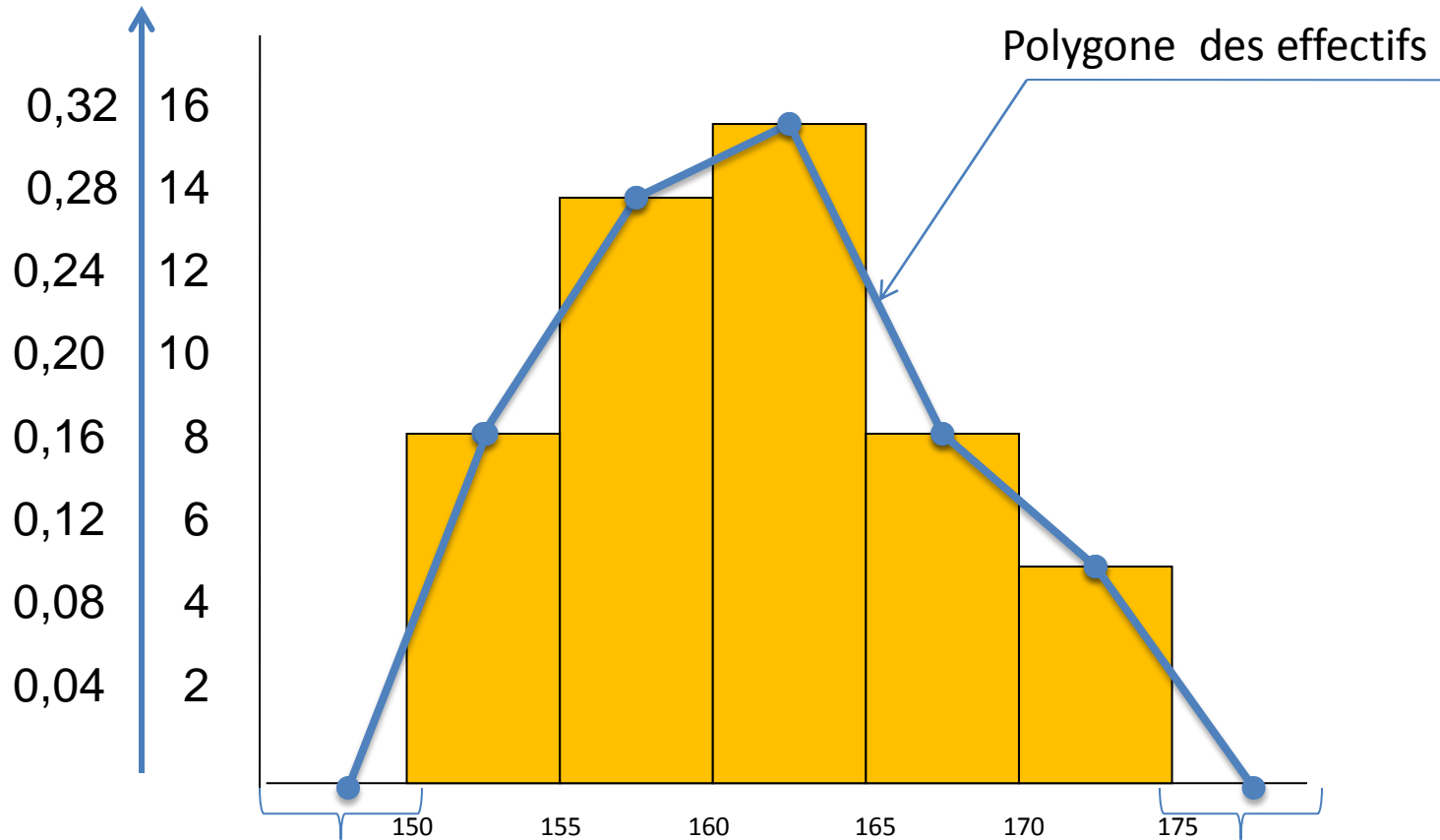
Modalités	Amplitude $A_i$	Effectifs $n_i$	ECC $N_i$	ECD $N_i$	Fréquences $f_i$	FCC $F_i$	FCD $F_i$
[ 150,155[	5	8	8	50	0,16	0,16	1
[155,160[	5	14	22	42	0,28	0,44	0,84
[160,165[	5	15	37	28	0,3	<b>0,74</b>	<b>0,56</b>
[165,170[	5	8	45	13	0,16	0,90	0,26
[170,175[	5	5	50	5	0,1	1	0,1
TOT		50			1		

0,74: On dit que 74 % des étudiants de la section ayant une taille moins de 165 cm.

0,56: On dit que 56 % des étudiants de la section ayant au moins une taille égale à 160 cm.

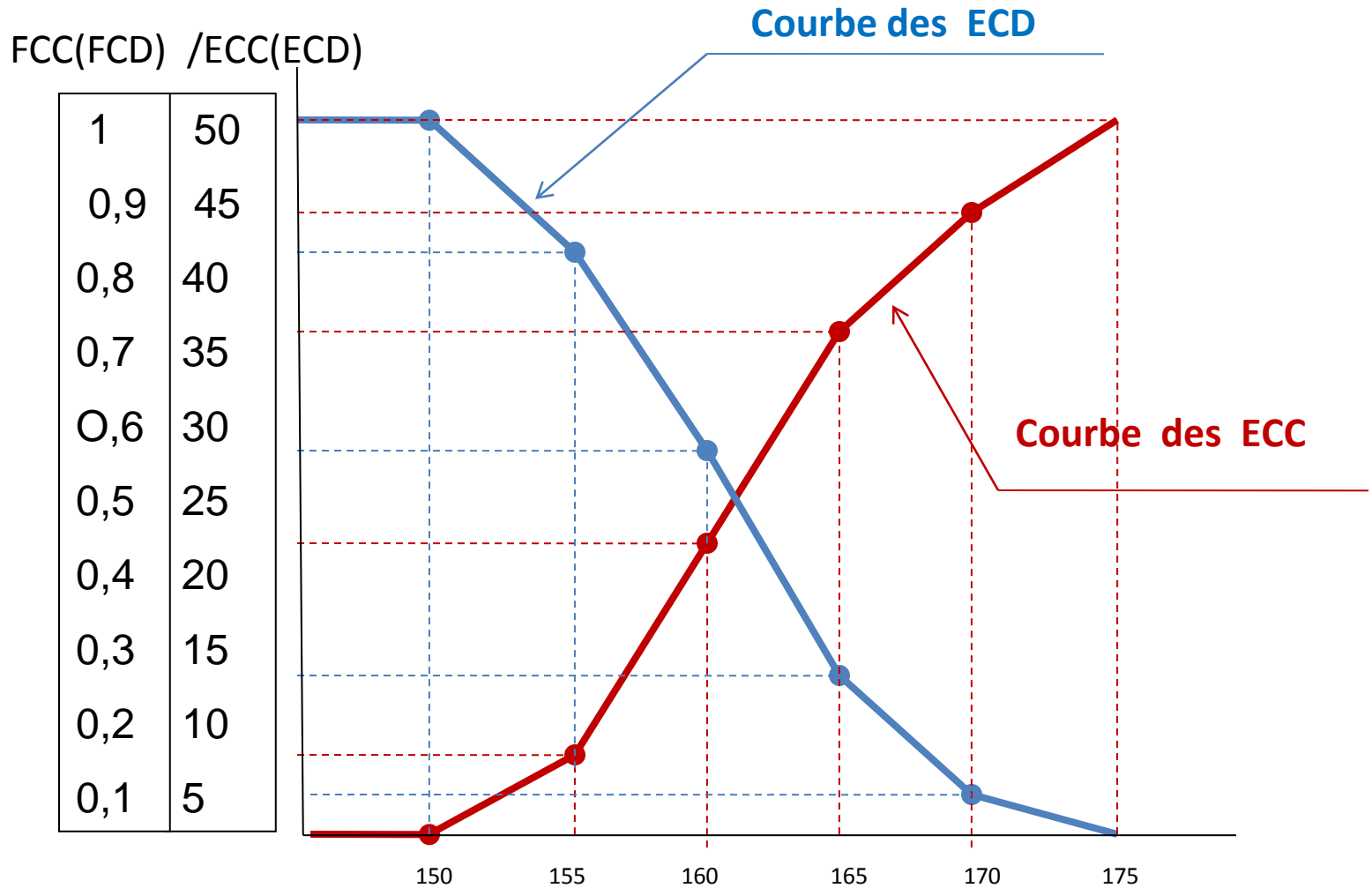
# B-Représentations graphiques

## cas des amplitudes égales





# C- Courbes es des effectifs cumulés croissants et décroissants



# D-tableau statistique

## 2<sup>ème</sup> cas : les amplitudes sont inégales

- Si au moins une classe a une amplitude différente des autres
  - on choisi une amplitude de référence :  $A_r$  (la plus petite ou la plus répandue)
  - puis on corrige la fréquence des classes différentes en la divisant par l'amplitude associée et en multipliant par l'amplitude de référence  $A_r$ .

# D-tableau statistique

2<sup>ème</sup> cas : les amplitudes sont inégales

*fréquence corrigée est donc :  $f'_i = f_i \times \left(\frac{A_r}{A_i}\right)$*

*avec  $f_i$  fréquence de la modalité  $x_i$*

*$A_r$  : amplitude de référence*

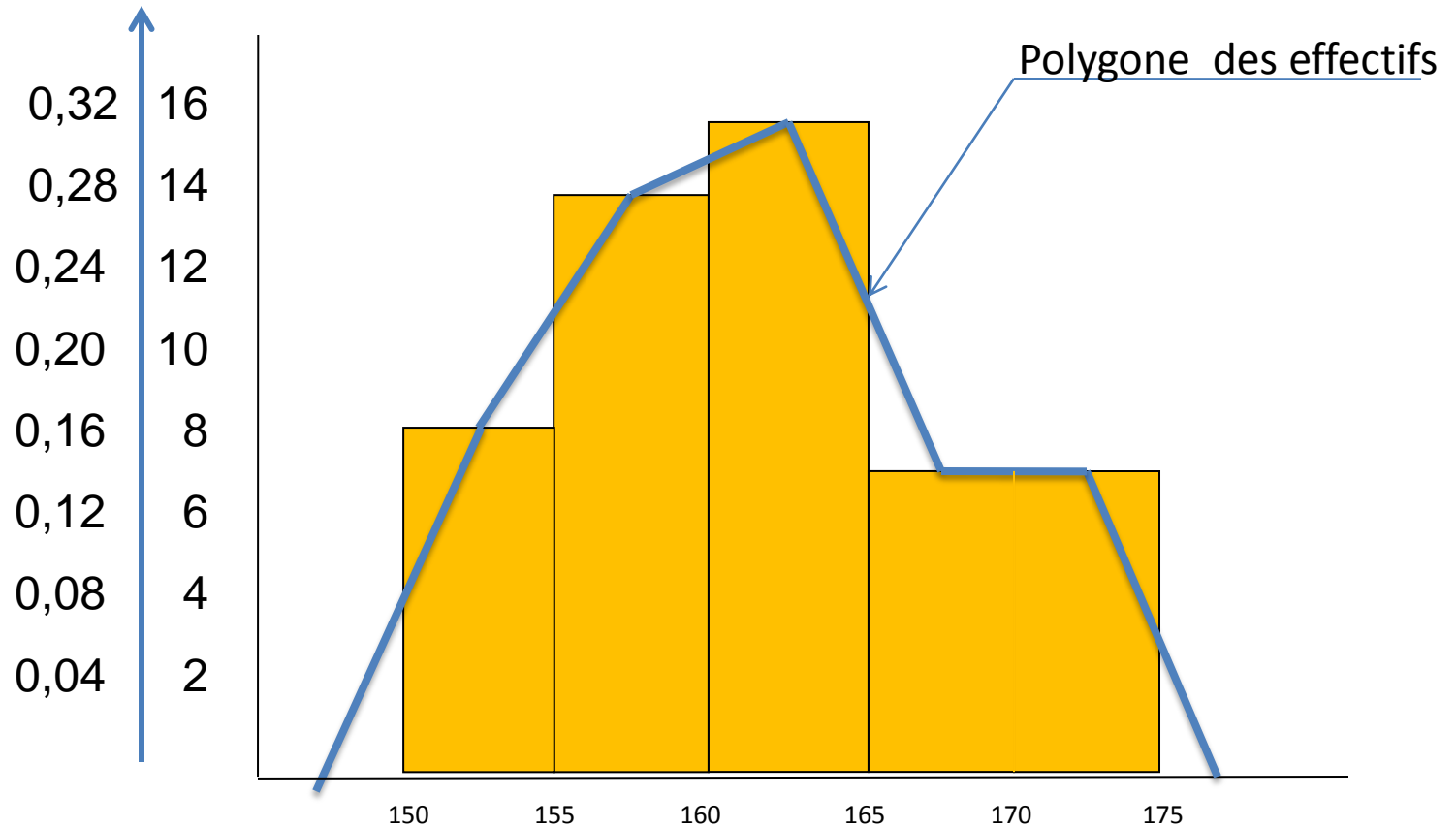
*et  $A_i$  : l'amplitude de la classe  $i$  ou de la modalité  $x_i$*

## A- Le tableau statistique

Exemple : Reprenons l'exemple de la taille en centimètres de 50 élèves d'une classe en regroupant les deux derniers intervalles

Modalités $X_i$	Effectifs $n_i$	Fréq $f_i$	Amplitude $A_i$	Effectifs corrigés $n_i'$	Fréquences corrigées $f_i'$
[ 150,155[	8	0,16	5	8	0,16
[155,160[	14	0,28	5	14	0,28
[160,165[	15	0,3	5	15	0,3
[165,175[	13	0,26	10	6,5	0,26 * (5/10)= 0,13
TOT	50	1			

## B- Représentations graphiques



# Application 2

On s'intéresse à la CSP des parents des étudiants de la section SC.ECO. Niveau 1<sup>ère</sup> année.

Nous avons interrogé 30 étudiants tirés au hasard parmi eux.

Le résultat de cette étude a été comme suit:

C Sup	Fonc	Fonc	PL	A C	C Sup	Sans	PL	Fonc	A C
Fonc	A C	Sans	PL	C Sup	Fonc	PL	PL	PL	C Sup
C Sup	Fonc	PL	Sans	Fonc	PL	C.Sup	Sans	AC	Fonc

- 1- Déterminer: la population, l'échantillon, l'individu, le caractère étudié et le type du caractère
- 2- Présenter cette distribution sous forme d'un tableau d'effectifs
- 3- Présenter graphiquement de deux façons différentes les données de ce tableau

# Application 2: Réponses

**la population** : les étudiants de la section SC.ECO. Niveau 1<sup>ère</sup> année.

**l'échantillon**: 30 étudiants tirés au hasard

**l'individu** : chaque étudiant de la population

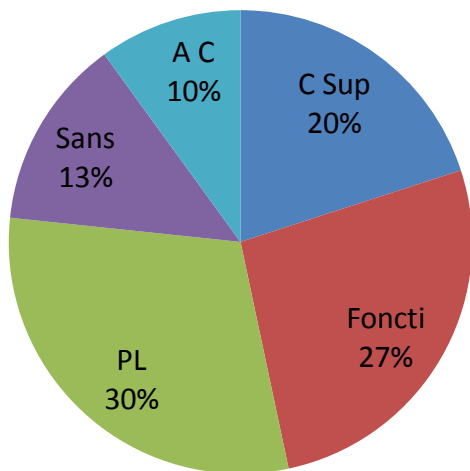
**le caractère étudié**: CSP des parents des étudiants .

**Type du caractère** : qualitatif nominal

## Le tableau statistique

Modalités	C Sup	Fonc	PL	Sans	A C	Tot
Effectifs	6	8	8	4	4	30
Fréquences	0,2	0,27	0,27	0,13	0,13	1

### Diagramme en secteurs de fréquences



### Diagramme en barres

