Statistiques

Chapitre 1 *: statistiques descriptives*

1- INTRODUCTION

La statistique est un ensemble de méthodes permettent de rassembler, organiser, analyser et interpréter des observations que l'on peut mesurer ou classer, dans un but de comparaison, de prévision, de constat. …etc.).

2- SERIES STATISTIQUES A UNE VARIABLE

Terminologie :

Population : Ensemble que l'on observe et qui sera soumis à une analyse statistique. Chaque élément de cet ensemble est un individu ou unité statistique.

Echantillon : est une partie de la population considérée.

Caractère (ou variable) : est une propriété étudiée sur chaque individu.

Les modalités du caractère : marié, célibataire, divorcé et veuf sont les modalités de l'etat matrimonial, par exemple.

Différents types de variables statistiques (caractère) :

- Lorsque la variable ne se prête pas à des valeurs numériques, elle est dite **qualitative(ou non mesurable)** exemple : Couleur des yeux – Nationalités-Groupe sanguin…etc).
- Lorsque la variable peut être exprimée numériquement, elle est dite **quantitative** (ou mesurable).Dans ce cas elle peut être discrète ou continue.
  - ✓ Elle est discrète :si elle ne prend que des valeurs isolées .Exemple : le nombre d'enfants d'une famille .
  - ✓ Dans le cas continu, le nombre de ces valeurs est très grand et on regroupe toutes ces valeurs en classes. (**Exp** :tailles, poids, salaires…etc).

A chaque valeur (classe) est associés à un effectif.

Effectif : L'effectif d'une classe ou d'une modalité est le nombre d'individu de cette classe ou de cette modalité. Généralement on note **ni.**

Effectif total : est la somme des effectifs de toutes les classes, on le note souvent N, on a alors :

N=∑ni = n1+n2+n3+…..+nk

Fréquence:

- **La fréquence fi** de la classe I ou de la modalité xi est le rapport **ni /N**, la fréquence d'une classe est un nombre de l'intervalle [0 ;1].
- **L'effectif cumulé** d'une modalité est la somme des effectifs des modalités qui lui sont inférieures ou égale $n_{cum} = \sum_{i=1}^{k} ni$.
- **La fréquence cumulée** d'une modalité est la somme des fréquences des modalités qui lui sont inférieures ou égales.

3- DISTRIBUTION D'UNE VARIABLE STATISTIQUE

On considère une série statistique X à caractère quantitatif, dont les k valeurs sont données par x1,x2,x3,….,xk avec n1+n2+n3+…..+nk =N

Xi = x1,x2,x3,….,xk ou bien Xi =(i= 1,2,3,…,k)modalités.

K : nombre de modalité

ni : Effectifs : nombre de fois que x répète une modalité.

N : Effectif total.

A chaque valeur (ou modalité) est associé une fréquence fi.

Fi : c'est la proportion d'individus associés à cette valeur fi=ni/N avec $\sum_{i=1}^{k} \boldsymbol{fi} = \mathbf{1}$.

Tableau Statistique

Modalités	Effectifs ni	Fréquences fi
M1	n1	f1
M2	n2	f2
M3	n3	f3
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.Mk	nk	fk
Total	N	1

4- SERIE STATISTIQUE ET REPRESENTATION GRAPHIQUE

4-1 Cas d'un caractère qualitatif :

On regroupe les résultats selon les modalités, le nombre d'individu pour chaque modalité est appelé effectif ou fréquence absolue, on note ni.

la fréquence relative fi=ni/N

Représentation graphique :

1. ***Un diagramme circulaire (sectoriel)***

   La mesure de chaque secteur angulaire est proportionnelle à l'effectif associé.

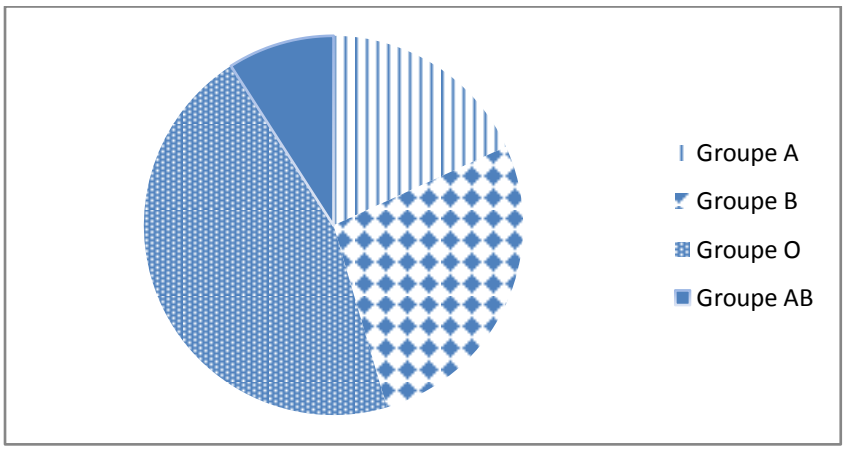


Figure : Diagramme Circulaire

2. ***Un diagramme en bandes :***

   Chaque classe est représentée par un rectangle de même largeur et de longueur proportionnelle à l'effectif, donc à la fréquence.

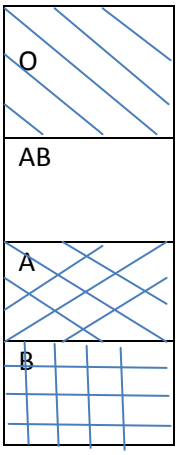


Figure : Diagramme en bande des effectifs

3.Diagramme en barre(en tuyaux d'orgue) :

Chaque classe est représentée par un rectangle de même largeur et de longueur proportionnelle à l'effectif donc à la fréquence.

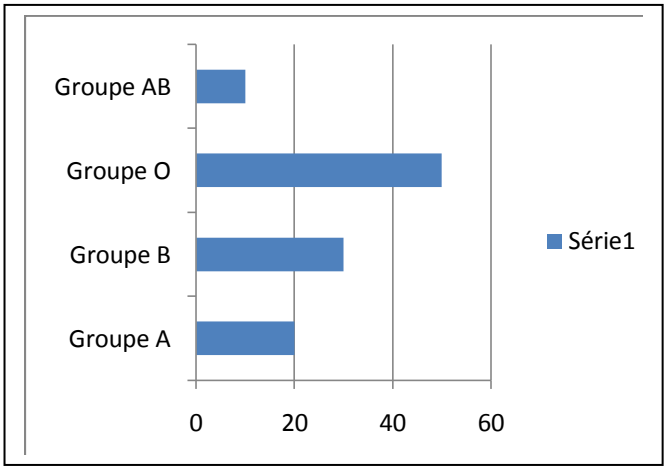


Diagramme en barre des effectifs

4-2 Cas d'un caractère quantitatif

4-2-1 Cas d'une variable discrète

On appelle variable quantitative discrète une variable quantitative ne prenant que des valeurs entières.

Exemple :

- le nombre d'enfants dans une population de favorite
- Le nombre d'année d'étude après le BAC dans une population d'étudiants ….etc.

a) ***Tableau Statistique***

Tableau statistique avec valeurs observées effectifs et fréquences

Modalités	Effectifs ni	Fréquences fi
x1	n1	f1
x2	n2	f2
x3	n3	f3
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.
.xk	nk	fk
Total	N	1

Exemple :

Notes obtenues à un contrôle dans une classe de 20 étudiants :

3 6 6 5 13 13 13 10 15 15 10 11 11 13 13 6 10 10 10 10

Notes	3	5	6	10	11	13	15	Total
ni	1	1	3	6	2	5	2	20
fi	0.05	0.05	0.15	0.3	0.1	0.25	0.1	1

b) ***Représentation graphique***

1) Diagramme en Bâtons :

Lorsque le caractère étudié est quantitatif discret, on peut représenter la série statistique étudiée par diagramme en bâtons : la hauteur de chaque bâton est alors proportionnelle à l'effectif (ou à la fréquence).associé à chaque valeur.

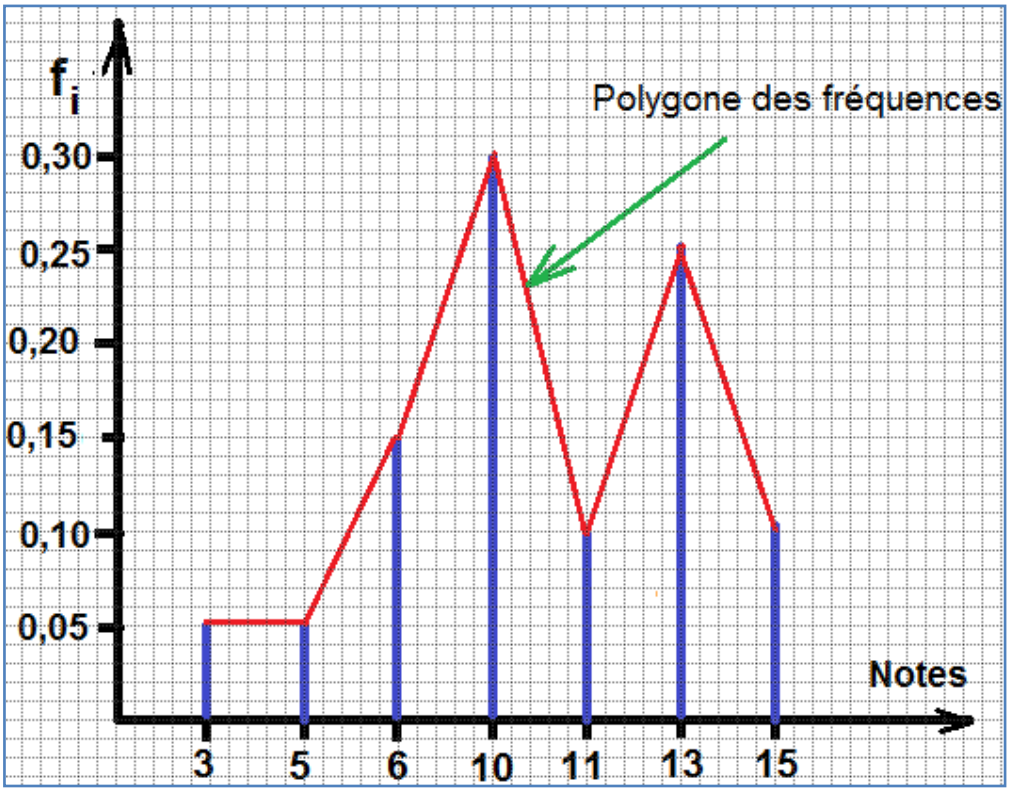


Figure : Diagramme en bâtons

4-2-2 Cas d'une variable continue

On appelle variable quantitative continue une variable qui peut prendre toutes les valeurs d'un intervalle fini ou infini ([a-b] ou [a-b [ ) avec : **b-a** : Amplitude de la classe, **(a+b)/2** : centre de la classe ***Exp :*** tailles, poids,…, etc.

a) ***Tableau Statistique***

Tableau statistique avec valeurs observées effectifs et fréquences

Modalités	Effectifs ni	Fréquences fi
[e1-e2[	n1	f1
[e2-e3[	n2	f2
[e3-e4[	n3	f3
.	.	.
.	.	.
.	.	.
.	.	.
..xk	.nk	fk
Total	N	1

Exemple

Les notes du module Maths 1 observées sur un échantillon de 100 étudiants sont représentées dans le tableau suivant :

Notes	[0-4[	[4-8[	[8-12[	[12-16[	[16-20[	Total
ni	5	20	30	35	10	100
amplitude	4	4	4	4	4	/
fi	0.05	0.2	0.3	0.35	0.1	1

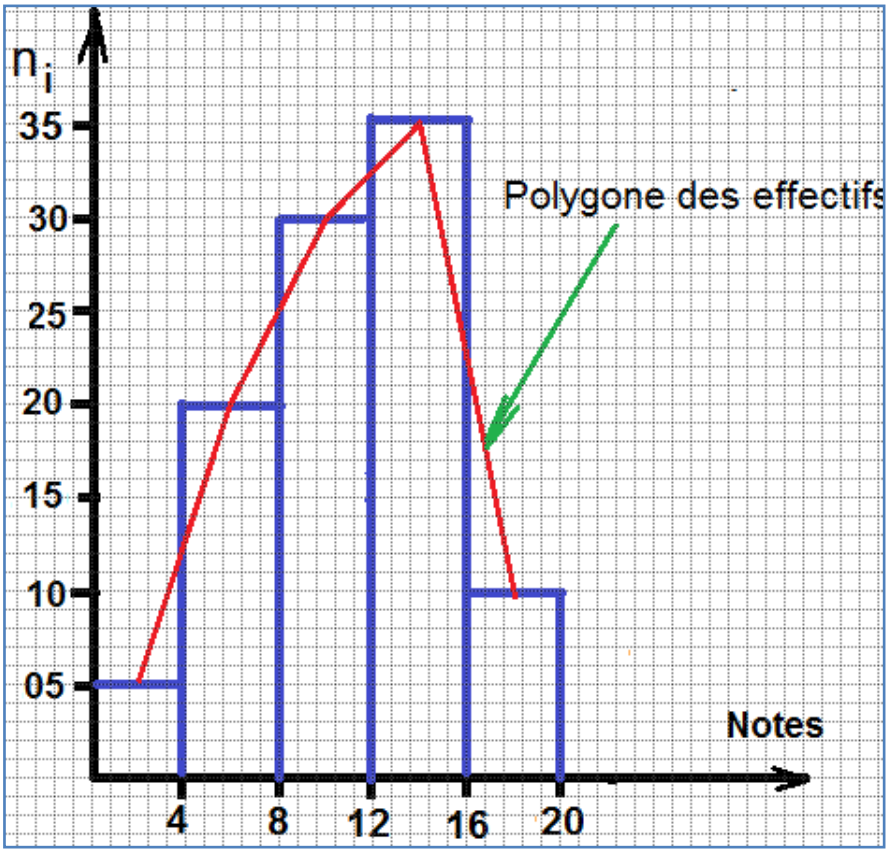
b) ***La Représentation graphique :***

Lorsque le caractère étudié est quantitatif continu, et lorsque les modalités sont regroupées en classe, on peut représenter la série statistique par un HISTOGRAMME : l'aire de chaque rectangle est alors proportionnelle à l'effectif (ou à la fréquence) associée à chaque classe.

Lorsque les classes ont la même Amplitude c'est la hauteur qui est proportionnelle à l'effectif.

- ***Même Amplitude***

Notes	[0-4[	[4-8[	[8-12[	[12-16[	[16-20[	Total
ni	5	20	30	35	10	100
amplitude	4	4	4	4	4	/
fi	0.05	0.2	0.3	0.35	0.1	1



Histogramme des effectifs

- ***Cas des Amplitudes différentes***

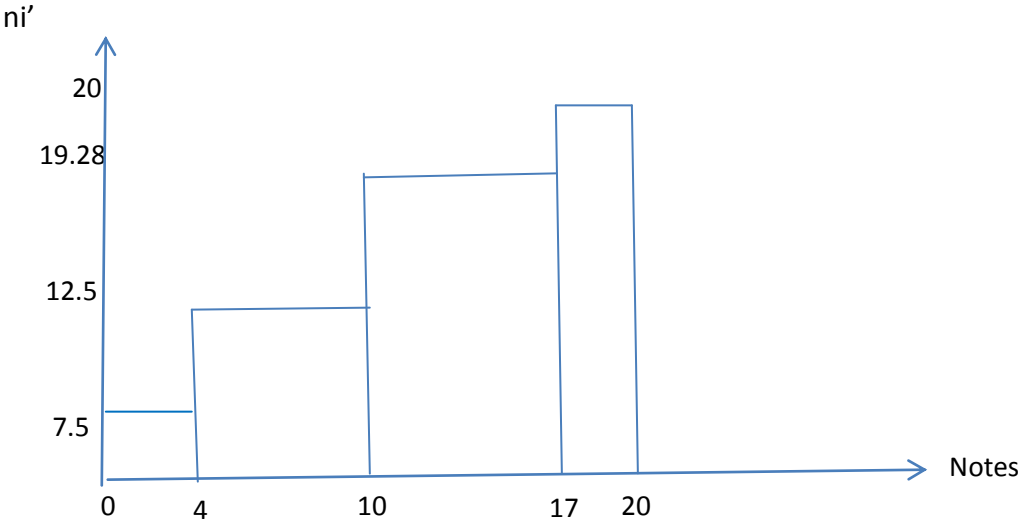
Pour tracer l'histogramme on utilise la méthode des effectifs rectifiés

ni' = ni *[la plus petite amplitude / amplitude de la classe] avec i=1,2,3,…,k

Exemple :

On reprend l'exemple précédent avec une nouvelle distribution.

Notes	[0-4[	[4-10[	[10-17[	[17-20[	Total
ni	10	25	45	20	100
amplitude	4	6	7	3	/
ni'	7.5	12.5	19.28	20	/



Histogramme des effectifs rectifiés

4-3 La courbe cumulative

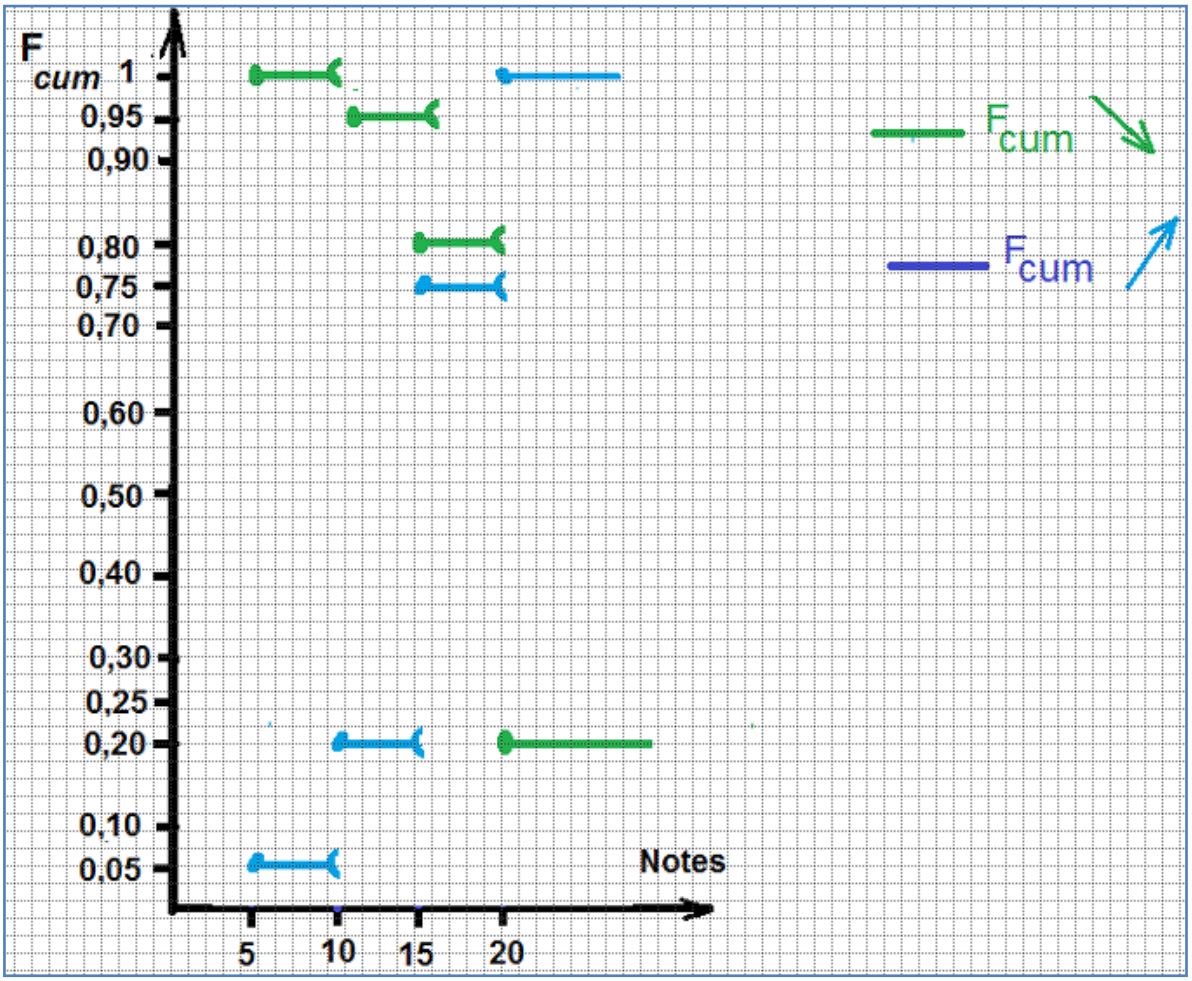
- ***Quantitatif Discret :***

$Fi_{cum}\nearrow = \sum_{i=1}^{k} fi = f1+f2+f3+....+fk$

$Fi_{cum}\searrow = 1 - \sum_{i=1}^{k-1} fi$

Exemple

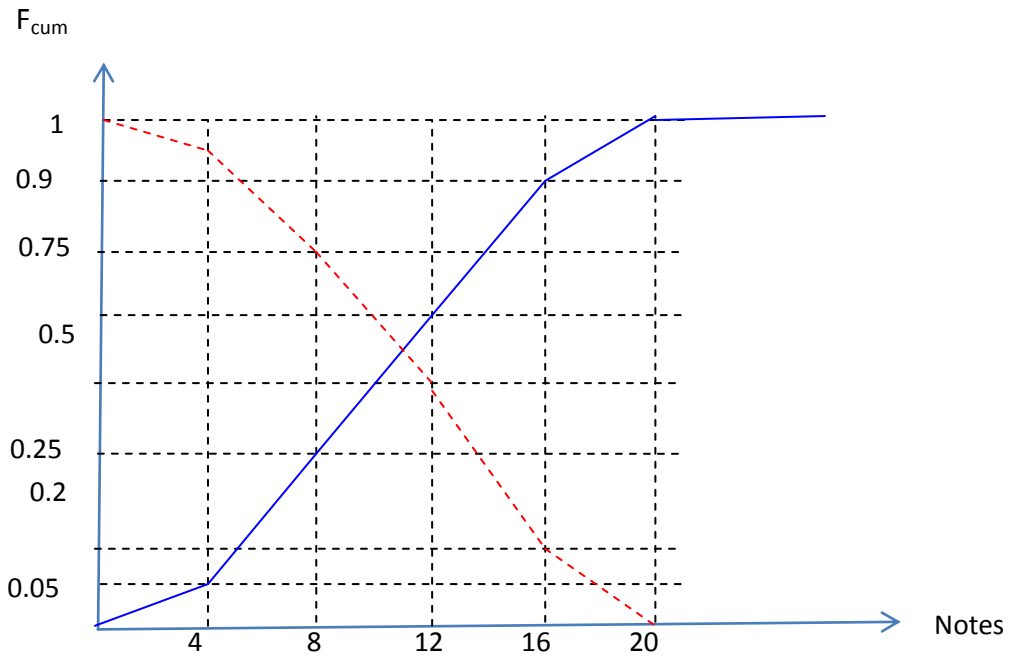
Notes	5	10	15	20	Total
ni	1	3	11	5	20
fi	0.05	0.15	0.55	0.25	1
$F_{cum}\nearrow$	0.05	0.2	0.75	1	/
$F_{cum}\searrow$	1	0.95	0.8	0.25	/



La courbe cumulative (la courbe d'Escalier)

Quantitatif Continu:

Notes	[0-4[	[4-8[	[8-12[	[12-16[	[16-20[	Total
ni	5	20	30	35	10	100
amplitude	4	4	4	4	4	/
fi	0.05	0.2	0.3	0.35	0.1	1
F_{cum} ↗	0.05	0.25	0.55	0.9	1	/
F_{cum} ↘	1	0.95	0.75	0.45	0.1	/



la courbe cumulative

4-4 Règle de STURGES [*procédure de regroupement en classes*]

On applique cette règle pour la transformation des données discrètes en données continues.

1. Ordonner les valeurs Xi de la série statistique à l'ordre croissant.
2. Calculer le nombre des classes, on le note par K avec K= 1+3.3logN .
3. Calculer l'étendu (e=Xmax -Xmin)
4. L'amplitude **a=e/K.**
5. Constitution des classes.[la classe est un intervalle fermé à gauche et ouvert à droite].

Exemple :

On a relevé les nombres d'allumettes continues respectivement dans 20 boites lors d'un contrôle dans une usine de fabrication ; les résultats sont les suivants :

40, 42, 32, 38, 40, 48, 30, 38, 36, 36, 40, 34, 40, 34, 40, 38, 40, 42, 44, 42

5 -Paramètres Caractéristiques :

5.1 Caractéristiques de position

5-1-1: Pour le cas discret

1- **Le mode** : le mode d'une variable statistique discrète X est la valeur xi qui correspond à l'effectif le plus élevé (la valeur la plus fréquente, dominante).

2- **La médiane** : la médiane est un paramètre de position, qui permet de couper la population étudiée en deux groupes contenant le même nombre d'individus. Ce paramètre est utile pour donner la répartition du caractère étudié, car 50% environ de la population étudiée a une modalité inferieur à la médiane et 50% une modalité supérieure à la médiane.

2-1Détermination de la médiane : (cas discret).

a) *Analytiquement :*

- *1er cas* : Si le nombre d'observation N = $\sum_{i=1}^{k} ni$ est ***impair***, dans ce cas la médiane est :

  $M_e = X_{\left[\frac{(N+1)}{2}\right]}$, où $X_{[(N+1/2)]}$ est l'observation numéro ((N+1)/2)dans l'ordre croissant.

- *2eme cas :* Si le nombre de l'observation $N = \sum_{i=1}^{k} ni$ est ***pair*** dans ce cas la médiane est : $M_e = \frac{1}{2}\left[X_{\left[\frac{N}{2}\right]} + X_{\left[\left(\frac{N}{2}\right)+1\right]}\right]$

b) *Graphiquement :*

1- On calcul F cum croissante.

2- La 1ere valeur qui correspond à F cum croissante égale à 0.5 nous donne la médiane .

3- ***Quartiles*** : ils partagent les observations en quatre parties égales.

Graphiquement :

Le 1er quartile noté Q1 tel que Fcum (Q1)=0.25.

Le 2eme quartile noté Q2 tel que Fcum (Q2)=0.5.

Le 3eme quartile noté Q3 tel que Fcum (Q3)=0.75.

Analytiquement :

Si N/4= nombre entier $\begin{cases} Q_1 = X_{(N/4)} \\ Q_2 = M_e = X_{\left(\frac{N}{2}\right)} \\ Q_3 = X_{(3N/4)} \end{cases}$

Si N/4 = nombre qui n'est pas entier, on prend la valeur supérieur directe .

4- ***La moyenne Arithmétique*** :

Soit une variable statistique dont les différentes modalités X1,X2,….,Xk aux quelles correspondent respectivement les effectifs n1,n2,n3,….nk avec n1+n2+n3+…+nk = N

La moyenne Arithmétique est donnée par :

$$\bar{X} = \left(\frac{1}{N}\right)\sum_{i=1}^{k} n_i x_i$$

Et comme *f*i = ni/N alors :

$$\bar{X} = \sum_{i=1}^{k} f_i x_i$$