Chapitre 2 : Séries statistiques à une variable :

1. **Effectif ou Fréquence absolue** :

(noté n_i) nombre d'apparitions de la valeur associé à un caractère dans un échantillon. ($n = \sum_{i=1}^{k} n_i$)

- Effectif cumulé croissant : (noté $N_i \nearrow$) somme cumulée du effectif avec tous ses effectifs précédents. $N_i \nearrow = \sum_{j=1}^{i} n_j$; $j \leq i$
- Effectif cumulé décroissant : (noté $N_i \searrow$) est donné par : $N_i \searrow = n - \sum_{j=1}^{i-1} n_j$

2. **Fréquence relative** :

(noté f_i) est le rapport de cet effectif à l'effectif total de la population. $f_i = \frac{n_i}{n}$; $i = 1, 2, \ldots, k$

- Fréquence cumulé croissante : (noté $F_i \nearrow$) $F_i \nearrow = \sum_{j=1}^{i} f_j$; $j \leq i$
- Fréquence cumulé croissante : (noté $F_i \searrow$) $F_i \searrow = 1 - \sum_{j=1}^{i-1} f_j$

Remarque:

- $0 \leq f_i \leq 1$
- f_i peut être exprimée en pourcentage %. ($f_i \times 100$)
- $\sum_{i=1}^{k} f_i = 1$

On présente tout les résultats dans le tableau suivant, en écrivant les modalités numériques en ordre croissant.

X_i	n_i	$N_i \nearrow$	$N_i \searrow$	f_i	$F_i \nearrow$	$F_i \searrow$
X_1	n_1	n_1	N	f_1	f_1	1
X_2	n_2	$n_1 + n_2$	$n - n_1$	f_2	$f_1 + f_2$	$1 - f_1$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
X_k	n_k	N	n_k	f_k	1	f_k
Total	$\sum_{i=1}^{k} n_i = n$			$\sum_{i=1}^{k} f_i = 1$		

Tableau 1 : Tableau statistique

Exemple 1 :

En notant X le nombre d'enfants de 20 familles, on a obtenu le tableau

0	2	2	3	1	3	4	2	2	3
1	0	1	4	2	2	1	3	1	2

ou la série ordonnée

00111112222222333344;

qui donne le tableau statistique suivant

Dr. Ahlam GUIATNI

X_i	n_i	$N_i \nearrow$	$N_i \searrow$	f_i	$F_i \nearrow$	$F_i \searrow$
0	2	2	20	0.10	0.10	1
1	5	7	18	0.25	0.35	0.90
2	7	14	13	0.35	0.70	0.65
3	4	18	6	0.20	0.90	0.30
4	2	20	2	0.10	1	0.1
Total	20			1		

Tableau 2 : Tableau statistique d'exemple 1

- La population : les 20 famille
- Le Caractère étudier X : Le nombre d'enfants
- Le type de caractère : Quantitatif discret
- Les modalité du caractère : $X_i \in \{0, 1, 2, 3, 4\}$

3. Représentations tabulaires et graphiques :

3.1. Caractère qualitatif :

X_i	n_i	f_i	θ_i	p_i
X_1	n_1	f_1	$f_1 \times 360$	$f_1 \times 100$
X_2	n_2	f_2	$f_2 \times 360$	$f_2 \times 100$
. . .	. . .	. . .	. . .	. . .
X_k	n_k	f_k	$f_k \times 360$	$f_k \times 100$
Total	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$	360	100

Tableau 3 : Tableau statistique « Caractère qualitatif »

Lorsque le caractère étudié est qualitatif on utilise un diagramme à bandes ou diagramme à secteurs.

i. **Diagramme à bandes :** C'est un diagramme qui à chaque modalité de la variable associé un rectangle de base constante et dont la hauteur est proportionnelle à l'effectif.

ii. **Diagramme à secteurs :** Ce type de graphique est formé d'un cercle divisé en secteurs, chaque secteur représente une catégorie particulière. La surface de chacun des secteurs est donné en degré par : $\theta_i = f_i .360$

Dr. Ahlam GUIATNI

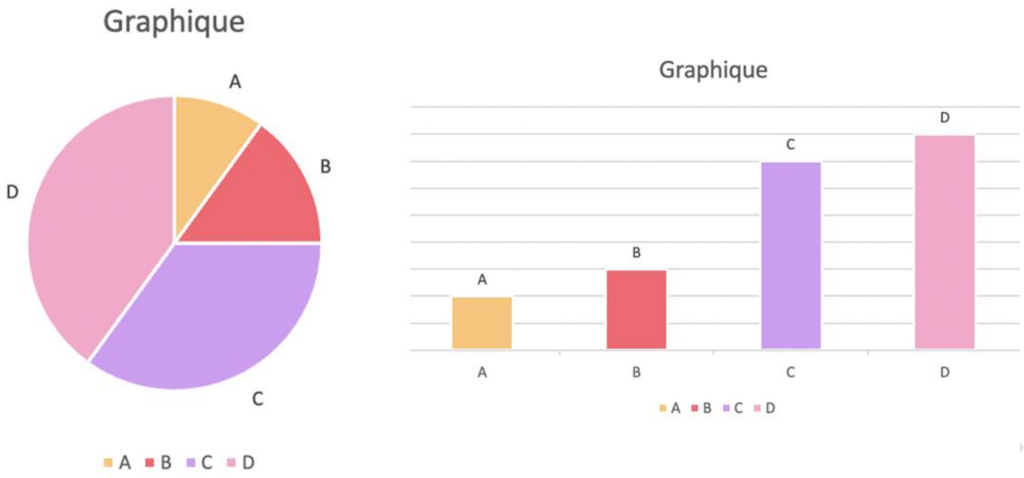

Figure 1:Diagramme à secteurs et diagramme à bandes

3.2. Caractère quantitatif discrète :

X_i	n_i	$N_i \nearrow$	$N_i \searrow$	f_i	$F_i \nearrow$	$F_i \searrow$
X_1	n_1	n_1	N	f_1	f_1	1
X_2	n_2	$n_1 + n_2$	$n - n_1$	f_2	$f_1 + f_2$	$1 - f_1$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
X_k	n_k	N	n_k	f_k	1	f_k
Total	$\sum_{i=1}^{k} n_i = n$			$\sum_{i=1}^{k} f_i = 1$		

Tableau 4 : Tableau statistique « Caractère quantitatif discret »

Lorsque le caractère étudié est quantitatif discret on utilise un diagramme en bâton.

- **Diagramme en bâton :** On associe un segment vertical dont la hauteur est proportionnelle à la valeur (effectif ou fréquence) connue.

Dr. Ahlam GUIATNI

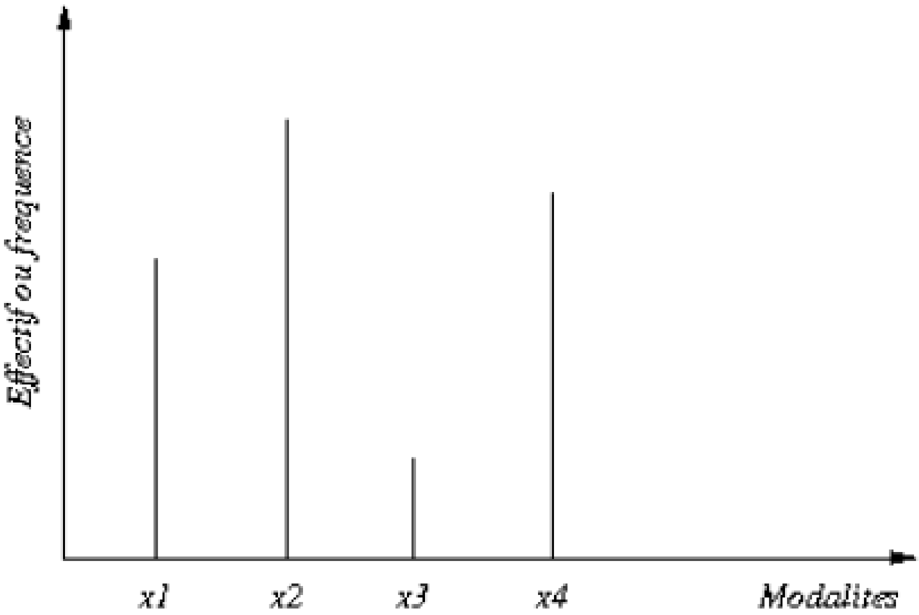

Figure 2: Diagramme en bâton

3.3. Caractère quantitatif continu :

Classes $[e_i , e_{i+1}[$	Centre c_i	n_i	$N_i \nearrow$	$N_i \searrow$	f_i	$F_i \nearrow$	$F_i \searrow$
$[e_1 , e_2[$	c_1	n_1	n_1	n	f_1	f_1	1
$[e_2 , e_3[$	c_2	n_2	$n_1 + n_2$	$n - n_1$	f_2	$f_1 + f_2$	$1 - f_1$
. . .		. . .	. . .	. . .	. . .	. . .	. . .
$[e_{k-1}, e_k[$	c_k	n_k	N	n_k	f_k	1	f_k
Total		$\sum_{i=1}^{k} n_i = n$			$\sum_{i=1}^{k} f_i = 1$		

Tableau 5: Tableau statistique « Caractère quantitatif continu »

- Une classe est un intervalle de type $[e_i, e_{i+1}[$
- Le centre de classe c_i est : $\frac{e_i + e_{i+1}}{2}$
- L'amplitude d'une classe est : $a_i = e_{i+1} - e_i$

Lorsque le caractère étudié est quantitatif continu on utilise un histogramme.

- **Histogramme :** C'est un diagramme composé des rectangles adjacents, chaque rectangle associé à chaque classe ayant une surface proportionnelle à l'effectif ou à la fréquence de cette classe.

Dr. Ahlam GUIA

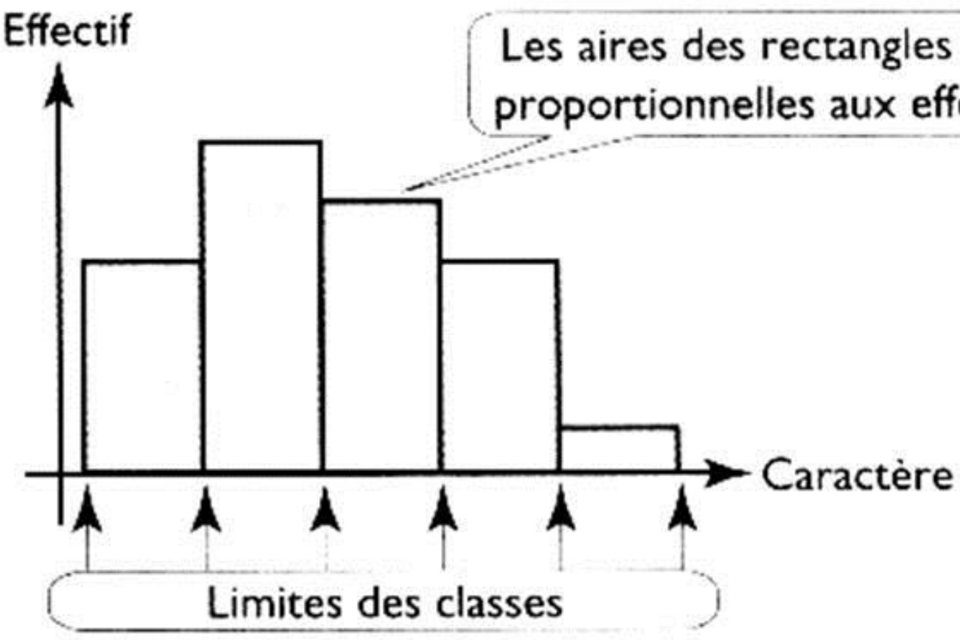

Figure 3 : Histogramme

Remarque :

Diagramme cumulatif est le diagramme représentatif de la fréquence cumulée ou d'effectif cumulé.

- Diagramme cumulatif pour une variable discrète : formé à l'aide d'un diagramme en escalier.

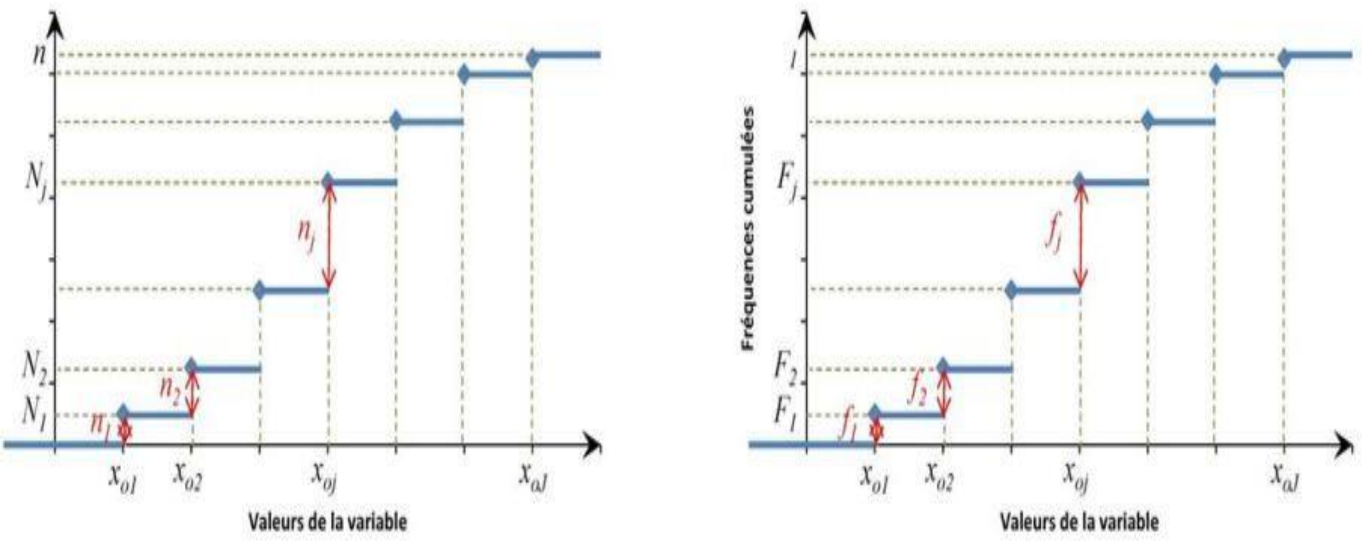

Figure 4: Diagramme en escalier

- Diagramme cumulatif pour une variable continue : formé à l'aide de segment de droite.

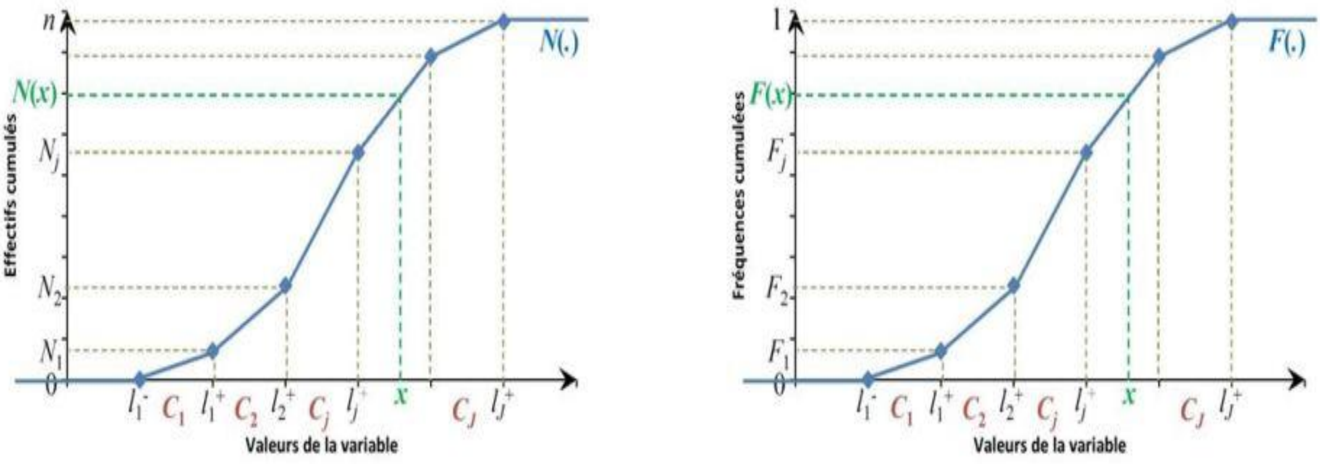

Figure 5 : Segment de droite

4. Caractéristiques de position :

Les paramètres de position ou « mesures de tendance centrale » sont des grandeurs susceptibles de représenter au mieux un ensemble de données. L'appellation «tendance centrale » vient du fait que ces paramètres donnent une idée de ce qui se passe au centre d'une distribution d'un ensemble de données. Les paramètres de position permettent de savoir autour de quelles valeurs se situent les valeurs d'une variable statistique.
On distingue trois mesures de tendance centrale :

4.1. Moyenne

La moyenne constitue l'un des paramètres fondamentaux de tendance centrale, la mesure la plus calculée et la plus utilisée lors de la description de séries statistiques mais non suffisant pour caractériser une distribution. Soit $\{x_1, x_2, \ldots, x_k\}$ une série statistique et $\{n_1, n_2, \ldots, n_k\}$ les effectifs correspondants.

4.1.1. Moyenne arithmétique : C'est la plus simple et la communément utilisée et ce, pas toujours à bon escient. Elle se note la plupart du temps par $\bar{x}$.

- Pour une variable discrète : $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$

- Pour une variable continue : $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i$

Exemple 2 : Soit la série statistique suivante : 12; 13; 4; 13; 12; 14; 15; 7; 15; 13; 12; 15; 7

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{3\times12+3\times13+1\times4+1\times14+3\times15+2\times7}{13} = \frac{152}{13} = 11.69$$

Exemple 3 : Soit la série statistique suivante :

Dr. Ahlam GUIATNI

Poids	[50, 55[	[55, 60[	[60, 65[	[65, 70[
n_i	2	4	8	6

On remplace les x_i par les centres de classes c_i :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \frac{2\times52.5+4\times57.5+8\times62.5+6\times67.5}{20} = \frac{1297.5}{20} = 64.875$$

4.1.2. Moyenne géométrique : La moyenne géométrique est un instrument permettant de calculer des taux moyens notamment des taux moyens annuels. La formule de la moyenne géométrique de cette série est donnée par : $\bar{G} = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$

4.1.3. Moyenne harmonique : On utilise la moyenne harmonique lorsqu'on veut déterminer un rapport moyen dans des domaines où il existe des liens de proportionnalité inverse. La formule de la moyenne harmonique de cette série est donnée par : $\bar{H} = \frac{n}{\sum_{i=1}^{k}\frac{n_i}{x_i}}$

4.1.4. Moyenne quadratique : Une moyenne qui trouve des applications lorsque l'on a affaire à des phénomènes présentant un caractère sinusoïdal avec alternance de valeurs positives et de valeurs négatives. Elle est, de ce fait, très utilisée en électricité. Elle permet notamment de calculer la grandeur d'un ensemble de nombre. A titre d'information. La formule de la moyenne quadratique de cette série est donnée par :
$\bar{Q} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2}$

4.2. Mode « Mo »:

- Pour une variable discrète : Le mode est la valeur de la sèrie ayant l'effectif le plus élevé $Mo = x_i|_{\max(n_i)}$. Lorsque la distribution a plus d'un mode, on parle d'une distribution «multimodale» (bimodale, trimodale, etc).
- Pour une variable continue : On cherche la classe modale abrégée [e_i, e_{i+1}[celle à effectifs le plus élevé. Pour avoir une valeur exacte, le mode se calcule de la manière suivante : $Mo = e_j + \frac{d_j}{d_{j+1} + d_j} \times a_j$

  e_j : limite inférieure de la classe modale ;
  a_j : amplitude de la classe modale ;
  d_j : écart d'effectif entre la classe modale et la classe inférieure la plus proche ;
  d_{j+1} : écart d'effectif entre la classe modale et la classe supérieure la plus proche.

Exemple 4 : Soit la série statistique suivante : 12; 13; 4; 13; 12; 14; 15; 7; 15; 12; 15; 7
$Mo_1 = 12$, $Mo_2 = 15$

Exemple 5 : Soit la série statistique suivante :

Dr. Ahlam GUIATNI

Poids	[50, 55[	[55, 60[	[60, 65[	[65, 70[	[70, 75[	[75, 80[	[80, 85[
n_i	2	5	12	16	14	8	3

La classe modale est [65, 70[

$e_j = 65$; $a_j = 5$; $d_j = 16\text{-}12$; $d_{j+1} = 16\text{-}14$

Alors, $Mo = 65 + \frac{4}{2+4} \times 5 = 68.33$

4.3. Médiane « Me »

Valeur centrale sur l'axe x divisant l'échantillon en 2 groupes égaux d'individus. Pour calculer Me, il faut d'abord ordonner la série.

- <u>Pour une variable discrète</u> : On désigne par n le nombre d'observations.
  - Si n est impair : Il est possible d'identifier simplement la valeur qui partage la population en deux effectifs égaux. Le rang central étant égal à $\frac{n+1}{2}$ alors $Me = x_{\frac{n+1}{2}}$.
  - Si n est pair : La médiane est alors égale à la moyenne des valeurs encadrant le milieu de la série alors $Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$
- <u>Pour une variable continue</u> : On cherche la classe contenant le $\frac{n}{2}$ individu de l'échantillon. Cette classe est appelée la classe médiane. En supposant que tous les individus de cette classe sont uniformément répartis à l'intérieur et $Me \in [e_j , e_{j+1}[$ la médiane se calcul de la façon suivante par interpolation linéaire :

$$Me = e_j + \frac{\frac{n}{2} - N_{j-1}\nearrow}{n_j} \times a_j$$

e_j : limite inférieure de la classe médiane ;
a_j : amplitude de la classe médiane ;
n_j : effectif de la classe médiane ;
$N_{j-1} \nearrow$: effectif cumulé inférieur à la classe médiane ;
n : taille de l'échantillon.

Exemple 6 : Soit la série statistique d'exemple 2 :
Le nombre d'effectif n = 13 impair, alors la médiane est la valeur d'ordre $\frac{n+1}{2} = \frac{13+1}{2} = 7$
La valeur d'ordre 7 est Me = 13.

Exemple 7 : Soit la série statistique d'exemple 4 :
Le nombre d'effectif n = 12 pair, alors la médiane est la moyenne arithmétique des valeurs d'ordre $\frac{n}{2} = \frac{12}{2} = 6$ et d'ordre $\frac{n}{2} + 1 = \frac{12}{2} + 1 = 7$

Dr. Ahlam GUIATNI

La valeur d'ordre 6 est 12

La valeur d'ordre 7 est 13; alors Me = $\frac{12+13}{2}$=12.5

Exemple 8 : Soit la série statistique d'exemple 5 :

On a $\frac{n}{2}$ = 30, la classe médiane est la classe qui leur effectif cumulé croissant supérieure ou égal 30, donc la classe médiane est : [65; 70[.

$e_j = 65$, $a_j = 5$, $n_j = 16$, $N_{j-1} \nearrow = 19$

Alors, Me = 65 + $\frac{30-19}{16} \times 5 = 68.44$

4.4. Quartiles, Déciles et Centiles

Les quartiles, déciles et centiles sont des caractéristiques qui correspondent au même genre de préoccupation que la médiane. Il s'agit des valeurs de la variable qui correspondent aux effectifs cumulés :

$\frac{n}{4}$, $\frac{2n}{4}$, $\frac{3n}{4}$ Pour les quartiles; le 2e quartile est la médiane.

$\frac{n}{10}, \frac{2n}{10}, \ldots, \frac{9n}{10}$ Pour les déciles; le 5e décile est la médiane.

$\frac{n}{100}, \frac{2n}{100}, \ldots, \frac{99n}{100}$ Pour les centiles; le 50e centile est la médiane.

On les appelle caractéristiques de position, puisqu'elles permettent de placer les valeurs de la variable.

Exemple 9 : Soit la série statistique suivante:

On relevé la taille en centimètres des joueurs d'une équipe de Basket : 203, 187, 185, 206, 180, 188, 198, 195, 200, 195, 218, 210.

On ordonne la série dans l'ordre croissant

x_i	180	185	187	188	195	198	200	203	206	210	218
n_i	1	1	1	1	2	1	1	1	1	1	1
$N_i \nearrow$	1	2	3	4	5	7	8	9	10	11	12

L'ordre de la valeur de Q_1 est : $\frac{1}{4} \times 12 = 3$; donc $Q_1 = 187$.

L'ordre de la valeur de Q_3 est : $\frac{3}{4} \times 12 = 9$; donc $Q_3 = 203$.

5. Caractéristiques de dispersion :

Les paramètres de dispersion nous renseignent sur la dispersion des valeurs autour de la valeur centrale de référence.

5.1. Etendue « E » :

Etendue d'une série statistique quantitative est la différence entre la plus grande valeur de x et la plus petite valeur. $E = x_{max} - x_{min}$

5.2. Variance « V(x) » :

Il est souvent intéressant de savoir si les valeurs sont très dispersées ou si elles sont proches de la moyenne. La variance est la caractéristique de dispersion la plus utilisée.

Dr. Ahlam GUIATNI

- <u>Pour un variable discrète</u> : $V(x) = \frac{1}{n} \sum_{i=1}^{k} n_i (x_i - \bar{x})^2$

- <u>Pour une variable continue</u> : $V(x) = \frac{1}{n} \sum_{i=1}^{k} n_i (c_i - \bar{x})^2$

5.3. Ecart-type « σ(x) » :

La racine carrée de la variance : $\sigma(x) = \sqrt{V(x)}$

5.4. Coefficient de variation « Cv » :

Si on veut comparer plusieurs séries statistiques ayant des moyennes très différentes, il vaut mieux se référer au coefficient de variation plutôt que l'écart-type. $Cv = \frac{\sigma_x}{\bar{x}} \times 100$

Remarque :

- Plus la valeur du coefficient de variation est faible, plus la dispersion autour de la moyenne est petite, plus la population est homogène.
- On considère qu'une distribution de données est homogène, lorsque Cv. est égal ou inférieur à 15%.

6. Caractéristiques de forme

Il existe deux mesures de forme qui caractérisent la forme des courbes représentant les distributions :
- Coefficient d'asymétrie
- Coefficient d'aplatissement

6.1. Coefficient d'asymétrie :

On se sert d'un coefficient pour mesurer l'asymétrie d'une distribution.
- Si le coefficient est nul (S_k = 0), alors il s'agit d'une distribution parfaitement symétrie. Et moyenne = médiane = mode

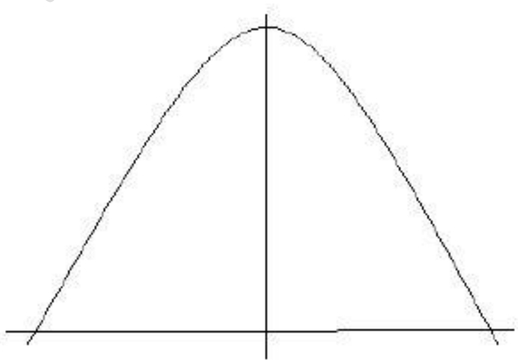

Figure 6 : Distribution parfaitement symétrie

Valeurs également distribués de part et d'autre de la valeur centrale.

- Si le coefficient est inférieur à 0, alors la distribution est du côté inférieur (étalée vers la gauche). Et moyenne < médiane < mode

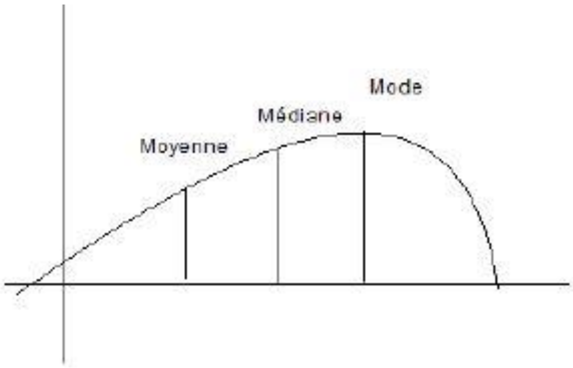

Figure 7 : Distribution étalée vers la gauche

La valeur la plus forte fréquence se situe à droite de la moyenne.

- Si le coefficient est supérieur à 0, alors la distribution est du côté supérieur (étalée vers la droite). Et moyenne > médiane > mode

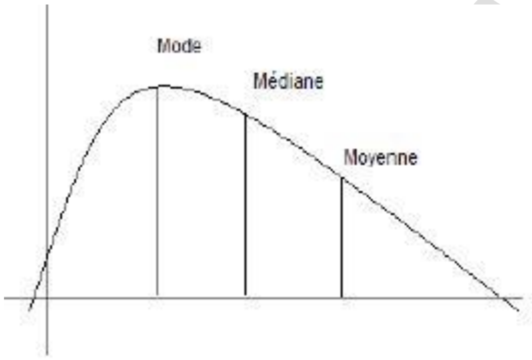

Figure 8 : Distribution étalée vers la droite

La valeur la plus forte fréquence se situe à gauche de la moyenne.

Calcul de coefficient :

- Coefficient d'asymétrie de Peason : On le note par S_p

$$S_p = \frac{3\,(\bar{x} - Me)}{\sigma_x} \text{ Ou } S_P = \frac{3\,(\bar{x} - Mo)}{\sigma_x}$$

- Si $S_P < 0$ distribution dissymétrique à gauche.
- Si $S_P = 0$ distribution symétrique.
- Si $S_P > 0$ distribution dissymétrique à droite.

- Coefficient d'asymétrie de Yule : On le note par S_Y

$$S_y = \frac{Q_3 + Q_1 - 2\,Q_2}{Q_3 - Q_1}$$

- Si $-1 \leq S_Y < 0$ distribution dissymétrique à gauche.
- Si $S_Y = 0$ distribution symétrique.
- Si $0 < S_Y \leq 1$ distribution dissymétrique à droite.

- Coefficient d'asymétrie de Fisher : On le note par S_F

$$S_F = \frac{M_3}{{\sigma_x}^3}$$

Ou $M_3 = \frac{1}{n} \sum n_i \,(x_i - \bar{x})^3$ est le moment centré d'ordre 3.
- Si $S_F < 0$ distribution dissymétrique à gauche.
- Si $S_F = 0$ distribution symétrique.
- Si $S_F > 0$ distribution dissymétrique à droite.

6.2. Coefficient d'aplatissement :

Pour mesurer l'aplatissement de la courbe, on utilise,

- Coefficient P_P de Peason : Le coefficient P_P de Peason basé sur le moment centré d'ordre 4.

$$P_P = \frac{M_4}{{\sigma_x}^4}$$

- $P_P > 3$ courbe leptocurtique ou hypernormale.
- $P_P = 0$ courbe normale.
- $P_P < 3$ courbe platicurtique ou hyponormale.

- Coefficient P_F de Fisher : Le coefficient P_F de Fisher est donné par

$$P_F = \frac{M_4}{{\sigma_x}^4} - 3$$

- $P_F > 0$ courbe leptocurtique ou hypernormale.
- $P_F = 0$ courbe normale.
- $P_F < 0$ courbe platicurtique ou hyponormale.

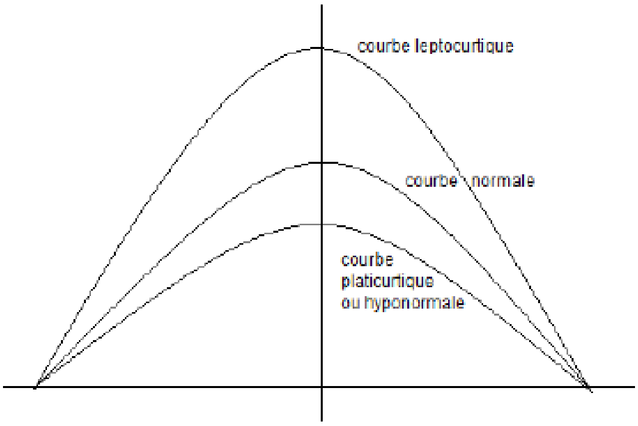

Figure 9 : Différentes courbes