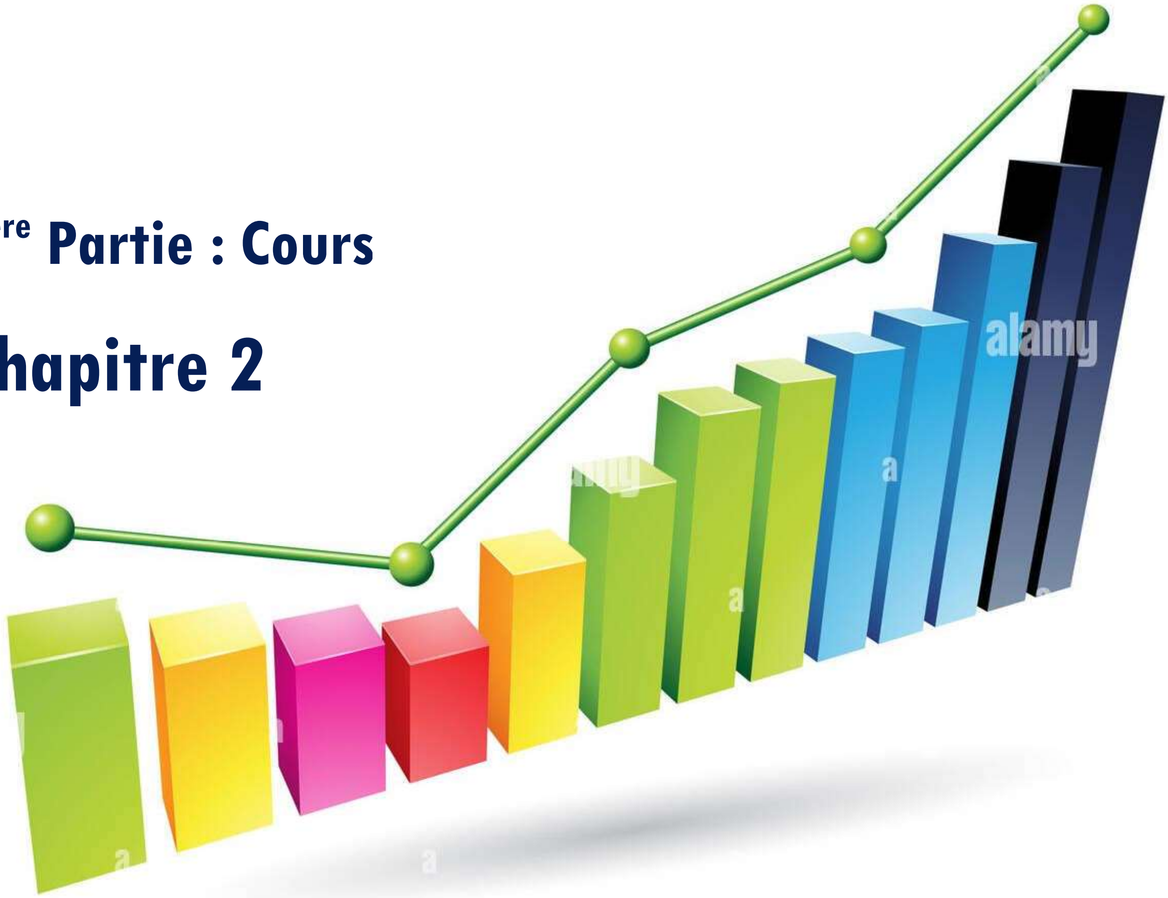
1ère Partie : Cours

Chapitre 2



Statistique descriptive univariée

Dans ce chapitre nous allons associer à chaque série statistique un ensemble de nombres réels qui sont appelés paramètres, d'une part des nombres qui se positionnent au milieu des valeurs dans un rangement croissant, d'autre part d'autres des nombres qui nous donnent la dispersion au tour d'un paramètre appelé moyenne arithmétique.

Ainsi on considère deux types de paramètres.

- Paramètres de positions
- Paramètres de dispersion.

2.1 Paramètres de position

2.1.1 Moyenne arithmétique

- Pour une variable quantitative discrète la moyenne arithmétique notée est donnée par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_rx_r}{N} = \sum_{i=1}^{i=r} \frac{n_ix_i}{N} = \sum_{i=1}^{i=r} f_ix_i$$

- Dans le cas d'une variable quantitative continue on a :

$$\bar{x} = \frac{n_1c_1 + n_2c_2 + \cdots + n_rc_r}{N} = \sum_{i=1}^{i=r} \frac{n_ic_i}{N} = \sum_{i=1}^{i=r} f_ic_i$$

avec $c_i = \frac{b_{i-1} + b_i}{2}$, le centre de la classe $[b_{i-1}, b_i[$.

Exemple 2.1.1
On reprend l'exemple de la page 10.

Classes	Centres c_i	Effectifs n_i	Fréquences f_i	$n_i\, c_i$
[20-40[	30	15	0,15	450
[40-60[	50	20	0,20	1000
[60-100[	80	20	0,20	1600
[100-200[	150	45	0,45	6750
Total		100	1,00	9800

$$\Rightarrow \bar{x} = \sum_{i=1}^{i=r} \frac{n_ic_i}{N} = \frac{9800}{100} = 98 \Rightarrow \bar{x} \in [60 - 100[$$

Remarque 2.1.1 La somme des écarts à la moyenne est nulle, c'est-à-dire

$$\sum_{i=1}^{i=r} n_i(x_i - \bar{x}) = 0$$

En effet

$$\sum_{i=1}^{i=r} n_i(x_i - \bar{x}) = \sum_{i=1}^{i=r} (n_ix_i - n_i\bar{x}) = \sum_{i=1}^{i=r} n_ix_i - \bar{x}\sum_{i=1}^{i=r} n_i = \bar{x}N - \bar{x}N = 0$$

2.1.2 Moyenne géométrique

Notée G et donnée par

$$G = \sqrt[N]{x_1^{n_1}.x_2^{n_2}\cdots x_r^{n_r}}$$

qui peut être exprimée en fonction des fréquences f_i :

$$G = \sqrt[N]{x_1^{n_1}.x_2^{n_2}\cdots x_r^{n_r}} = x_1^{\frac{n_1}{N}}.x_2^{\frac{n_2}{N}}\cdots x_r^{\frac{n_r}{N}} = x_1^{f_1}.x_2^{f_2}\cdots x_r^{f_r}$$

Remarque 2.1.2 La moyenne géométrique G vérifie

$$\ln G = \frac{1}{N}\sum_{i=1}^{i=r} n_i \ln x_i.$$

2.1.3 La moyenne harmonique

Notée H et donnée par

$$H=\frac{1}{\frac{1}{N}\sum_{i=1}^{i=r}\left(\frac{n_i}{x_i}\right)}=\frac{1}{\sum_{i=1}^{i=r}\left(\frac{f_i}{x_i}\right)}=\frac{N}{\sum_{i=1}^{i=r}\left(\frac{n_i}{x_i}\right)}$$

2.1.4 La moyenne quadratique

Notée Q et donnée par $Q=\sqrt{\frac{1}{N}\sum_{i=1}^{i=r}n_i x_i^{\;2}}$

Remarque 2.1.3 $\bar{x}$, G, H, Q vérifient

$$x_{\min}\leq H\leq G\leq \bar{x}\leq Q\leq x_{\max}$$

Exemple 2.1.2

Le calcul de $\bar{x}$, G, H, Q de la série : 2, 5, 11, 18.

$$\Rightarrow \bar{x}=\frac{2+5+11+18}{4}=9,\quad G=\sqrt[4]{2\times 5\times 11\times 18}=6{,}67,$$

$Q=\sqrt{\frac{1}{4}\left(2^2+5^2+11^2+18^2\right)}=10{,}88$ et $H=\dfrac{4}{\frac{1}{2}+\frac{1}{5}+\frac{1}{11}+\frac{1}{18}}=4{,}72$

Et on vérifie que

$$x_{\min}=2\leq H=4{,}72\leq G=6{,}67\leq \bar{x}=9\leq Q=10{,}88\leq x_{\max}=18$$

Remarque 2.1.4 Avec deux valeurs x_1, x_2 distinctes ($n_1=n_2=1$) on a :

$$G=G(x_1,x_2)=G(\bar{x},H)$$

En effet

$$H=\frac{1}{\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)}=\frac{2}{\left(\frac{x_1+x_2}{x_1x_2}\right)}=\frac{x_1x_2}{\left(\frac{x_1+x_2}{2}\right)}$$

$$\Rightarrow G^2=\bar{x}.H\Rightarrow G=\sqrt{\bar{x}.H}\quad\Rightarrow G(x_1,x_2)=G(\bar{x},H)$$

Remarque 2.1.5 Avec deux valeurs x_1, x_2 distinctes ($n_1=n_2=1$) si on pose :

$$y_1 = (H - x_1) \quad et \quad y_2 = (H - x_2)$$

on a

$$H = H(x_1, x_2) = H(y_1, y_2)$$

2.1.5 Le mode

C'est la valeur de la variable ayant le plus grand effectif (ou la fréquence la plus élevée). On note le mode M_o.

- **Cas quantitatif discret**

Exemple 2.1.3
On considère les notes obtenues en statistique par un groupe de 20 étudiants :

7, 13, 5, 15, 12, 9, 7, 8, 14, 16, 13, 6, 13, 10, 13, 12, 10, 7,12, 13.

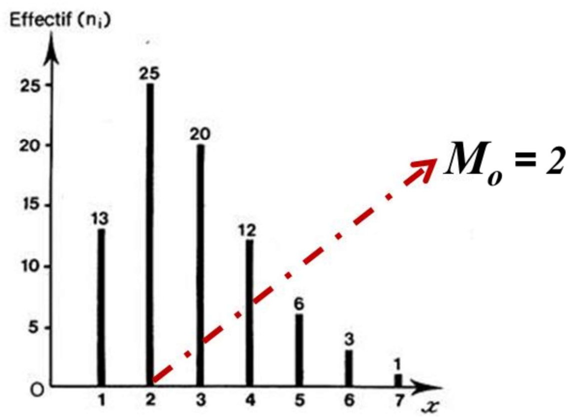
$\Rightarrow$

x_i	5	6	7	8	9	10	12	13	14	15	16
n_i	1	1	3	1	1	2	3	5	1	1	1

Le mode de cette série est $M_o = 13$, valeur qui apparaît cinq fois.
L'interprétation : Est que **la note la plus fréquente est 13**.

Remarque 2.1.6 Graphiquement, dans un diagramme en bâton le mode correspond à **l'abscisse du bâton le plus élevé**.

C'est-à-dire :



- **Cas quantitatif continu**

Dans ce cas, on parle plutôt de **classe modale.** On a deux cas.

1er cas : Cas d'amplitudes égales

Si les classes sont de même amplitude a_i (ie $a_i = a_j \, \forall \, i \neq j$), la classe modale est la classe d'effectif n_i le plus élevé, soit $[b_{i-1}, b_i[$, avec le mode $M_o \in [b_{i-1}, b_i[$ alors :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1+m_2}\right) \text{ avec}$$

b_{i-1} : la borne inférieure de la classe modale.

b_i : la borne supérieure de la classe modale.

a_i : l'amplitude de la classe modale.

$m_1 = n_i - n_{i-1}$ et $m_2 = n_i - n_{i+1}$.

Exemple 2.1.4

Soit la distribution de la population de 20 ménages selon le revenu (en centaines de DA) des deux parents

Classes en *CDA*	a_i	n_i	f_i
[200-300[	100	40	0,20
[300-400[	100	60	0,30
[400-500[	100	30	0,15
[500-600[	100	50	0,25
[600-700[	100	20	0,10
Total		200	1,00

Comme $n_2 = 60$ est le plus grand effectif, donc la classe modale est $[300 - 400[$. Le mode est calculé par :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1+m_2}\right) = 300 + 100\left(\frac{60-40}{(60-40)+(60-30)}\right)$$

$$\Rightarrow M_O = 340 \quad CDA$$

<u>Interprétation</u> : On dit que le salaire le plus fréquent est de 340 CDA.

<u>2</u>

2eme cas : Cas d'amplitudes inégales

Si les classes sont d'amplitude inégales a_i (ie $a_i \neq a_j$), **la classe modale** est la classe d'effectif corrigé n_i^c le plus élevé (ou encore la fréquence corrigée f_i^c la plus élevée), soit $[b_{i-1}, b_i[$, avec le Mode $M_o \in [b_{i-1}, b_i[$ est tel que :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1+m_2}\right) \text{ avec}$$

b_{i-1} : la borne inférieure de la classe modale.

b_i : la borne supérieure de la classe modale.

a_i : l'amplitude de la classe modale.

$m_1 = h_i - h_{i-1} = n_i^c - n_{i-1}^c$ et $m_2 = h_i - h_{i+1} = n_i^c - n_{i+1}^c$.

Où h_i, h_{i-1} et h_{i+1} sont les effectifs corrigés.

Remarque 2.1.7 Dans les 2 cas on peut calculer le mode M_o en utilisant les fréquences à la place des effectifs, en prenant

1. $m_1 = f_i - f_{i-1}$ et $m_2 = f_i - f_{i+1}$ si $(a_i = a_j \ \forall\ i \neq j)$
2. $m_1 = f_i^c - f_{i-1}^c$ et $m_2 = f_i^c - f_{i+1}^c$ si $(a_i \neq a_j)$.

Exemple 2.1.5

Soit la répartition de 100 personnes selon leur âge, on prend $a^* = 100$

Classes	a_i	n_i	d_i	n_i^c
[20 , 30[	10	20	2,00	200
[30 , 40[	10	25	2,50	250
[40 , 60[	20	35	1,75	175
[60 , 80[	20	20	1,00	100

Comme $n_2^c = 250$ est le plus grand effectif corrigé, donc la classe modale est [30, 40[. Le mode est calculé par :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right) = 30 + 10\left(\frac{250 - 200}{(250 - 200) + (250 - 175)}\right)$$

$$\Rightarrow M_O = 34.$$

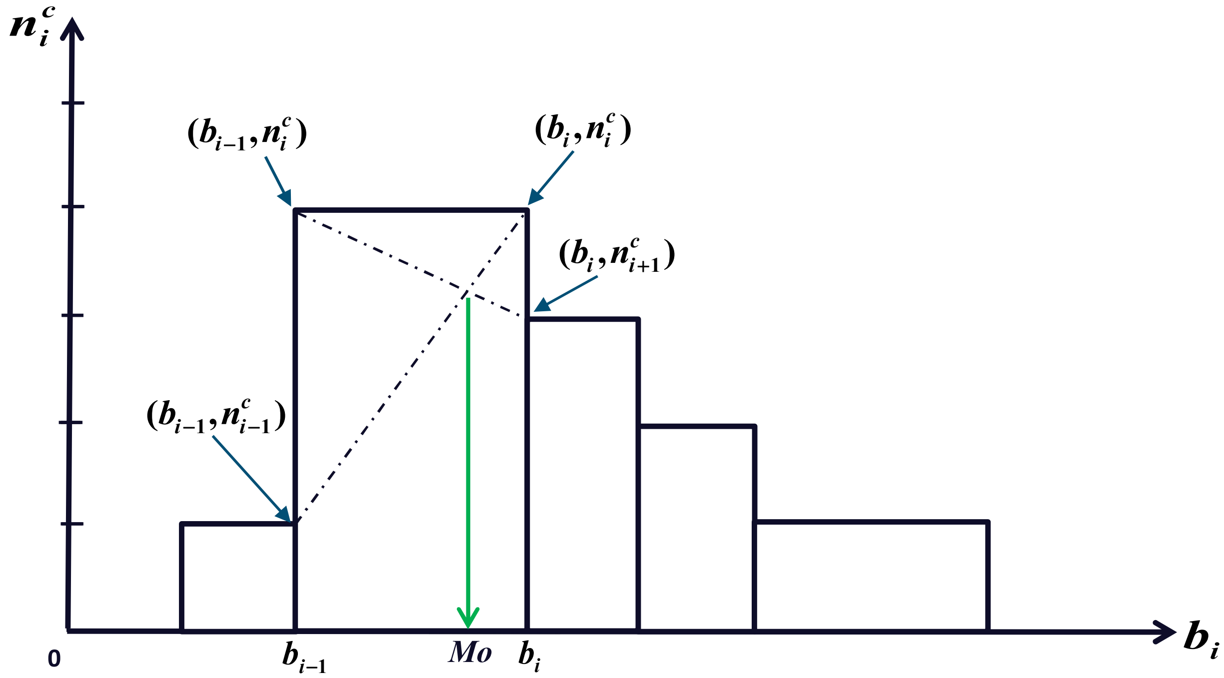
<u>**Interprétation**</u> : L'âge le plus fréquent est de 34 ans.

<u>Détermination graphique du mode</u>

Si **la classe modale** est $[b_{i-1}, b_i[$, avec le mode $M_o \in [b_{i-1}, b_i[$ alors :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right)$$

Et graphiquement le mode $M_o \in [b_{i-1}, b_i[$ sur un histogramme est le point d'intersection des deux segments $[(b_{i-1}, h_i); (b_i, h_{i+1})]$ et $[(b_{i-1}, h_{i-1}); (b_i, h_i)]$, voir figure suivante :



2.1.6 La médiane

Pour une série statistique rangée par ordre croissant la médiane $Mé$ est la valeur de la variable qui partage la population en deux groupes d'effectifs égaux.

- **Cas quantitatif discret**

Pour une série statistique rangée par ordre croissant c'est-à-dire :

$v_1 \le v_2 \le \cdots \le v_N$ la médiane $Mé$ est la valeur du milieu qui dépendra de l'effectif total N.

1. Si N est impair ($N = 2k+1$), alors $Mé = v_{k+1}$.
2. Si N est pair ($N = 2k$), alors $Mé = \dfrac{v_k + v_{k+1}}{2}$.

Exemples 2.1.6

1. Soit la répartition de 9 ménages selon le nombre d'enfants

x_i	0	1	2	3	4
n_i	2	2	1	3	1