

Module de Mathématiques Statistiques

Chapitre 2 : Statistiques Descriptives à une dimension

Séance 04

Responsables du cours: **Dr. Metiri Farouk & Dr. Sadoun
Ahmed & Dr. Aidi Khaoula**
Université de Badji Mokhtar -Annaba-
1^{ère} année -Département de TCSNV-
2021/2022

Mail address: fmetiri@yahoo.fr
saadounahmed1@yahoo.fr
khaoula.aidi@yahoo.fr

Etude d'une variable quantitative continue:

- Lorsque le caractère quantitatif comprend un grand nombre de valeurs, nous préférons regrouper les valeurs en intervalles appelés classes pour rendre la statistique plus lisible. Nous partageons alors l'ensemble des valeurs du caractère en classes $[a; b[$ avec $a < b$. On choisit les classes pas trop nombreuses, mais elles suffisaient pour qu'il n'y ait pas de perte d'information.

Exemple 1:

Lors d'une étude sur l'obésité, on a mesuré l'augmentation du poids (en **Pounds**, **1Kg = 2.22Pounds**) chez **20** personnes obèses. Voici les résultats obtenus:

Poids(Pound)	Effectif n_i
[10; 20[4
[20; 30[3
[30; 40[6
[40; 50[5
[50; 60[2
Total	N = 20

Représentation graphique

On représente une série statistique d'un caractère quantitatif continu par un **histogramme**:

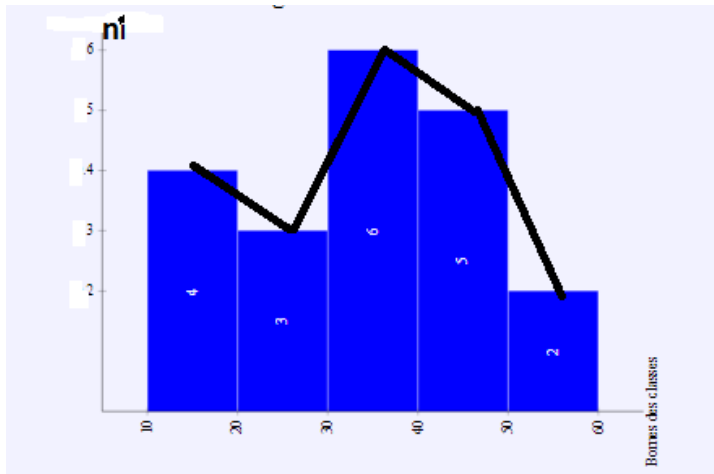


Figure: Histogramme et polygone des effectifs

courbes des effectifs cumulés croissants et décroissants :

On ajoute au tableau les 2 colonnes des $n_i^C \uparrow$ et $n_i^C \downarrow$

Poids	Effectif n_i	$n_i^C \uparrow$	$n_i^C \downarrow$
[10; 20[4	4	20
[20; 30[3	7	16
[30; 40[6	13	13
[40; 50[5	18	7
[50; 60[2	20	2
Total	$N = 20$		

Tracer le polygone des effectifs cumulés croissants:

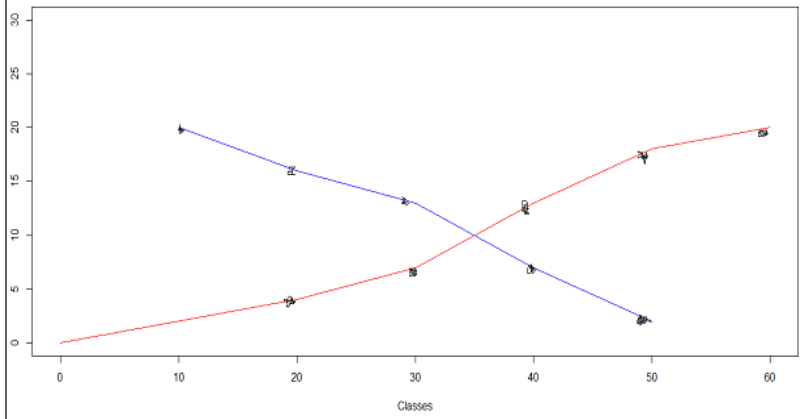
Pour tracer la courbe des effectifs cumulés croissants ECC :

- tracer un repère avec en abscisse les classes et en ordonnées les ECC
- placer chaque point dont l'abscisse est la borne **supérieure** de la classe et l'ordonnée le ECC associé à cette classe
- joindre tous les points par une ligne brisée, en ajoutant le point d'abscisse égale à la borne inférieure de la plus petite classe et d'ordonnée **0**

Pour tracer la courbe des effectifs cumulés décroissants ECD:

- tracer un repère avec en abscisse les classes et en ordonnées les ECD
- placer chaque point dont l'abscisse est la **borne inférieure** de la classe et l'ordonnée est le ECD associé à cette classe
- joindre tous les points par une ligne brisée, en ajoutant le point d'abscisse égale à la **borne supérieure de la plus grande classe et d'ordonnée 0.**

Courbes des effectifs cumulés croissants et décroissants



Les paramètres de position

1- La moyenne arithmétique

La moyenne arithmétique ou moyenne, qu'on désigne par \bar{x} est donnée par:

$$\bar{x} = \frac{1}{N} \sum_i n_i x_i$$

- Pour un caractère quantitatif continu, on doit ajouter une colonne au tableau statistique appelée:

$$\text{centre de classe} = x_i = \frac{a_i + b_i}{2}$$

Poids	Effectif n_i	centre	$n_i^C \uparrow$	$n_i^C \downarrow$
[10; 20[4	$\frac{10+20}{2} = 15$	4	20
[20; 30[3	25	7	16
[30; 40[6	35	13	13
[40; 50[5	45	18	7
[50; 60[2	55	20	2
Total	$N = 20$			

Pour l'exemple 1, La moyenne est

$$\bar{x} = \frac{(4 \times 15) + (3 \times 25) + (6 \times 35) + (5 \times 45) + (2 \times 55)}{20} = 34 \text{ Pounds}$$

2 – Le Mode :

Dans le cas des effectifs groupés par classes d'amplitudes égales, la classe modale est la classe d'effectif n_i le plus répété, soit $[a_i; b_i[$ avec $a_i < b_i$. L'effectif de la classe qui précède la classe modale est n_{i-1} et celui de la classe qui suit la classe modale est n_{i+1} alors:

$$M_O = a_i + \left[(b_i - a_i) \left(\frac{d_1}{d_1 + d_2} \right) \right]$$

avec

$$\begin{cases} d_1 = n_i - n_{i-1} \\ d_2 = n_i - n_{i+1} \end{cases}$$

Pour l'exemple 1, La classe modale qui possède le plus grand effectif est $[30; 40[$, donc $M_O \in [30; 40[$.

On a:

$$\begin{cases} d_1 = 6 - 3 = 3 \\ d = 6 - 5 = 1 \\ a_i = 30 \text{ et } b_i = 40 \\ b_i - a_i = 40 - 30 = 10 \end{cases}$$

Le mode est $M_O = 30 + \left[10 \left(\frac{3}{3+1} \right) \right] = 37,5$ Pounds.

3 – La médiane

Le calcul de la médiane dans le cas d'une variable continue passe d'abord par la détermination de la classe médiane, la classe où l'effectif cumulé croissant est supérieur à $\frac{N}{2}$, soit la classe $[a_i; b_i[$.

Poids	Effectif	centre	$n_i^C \uparrow$	$n_i^C \downarrow$
$[10; 20[$	4	$\frac{10+20}{2} = 15$	4	20
$[20; 30[$	3	25	7	16
$[30; 40[$	6	35	13	13
$[40; 50[$	5	45	18	7
$[50; 60[$	2	55	20	2
Total	$N = 20$			

Ensuite, par interpolation linéaire, on peut calculer la valeur précise de la médiane à l'intérieur de la classe médiane.

$$Me = a_i + \left[(b_i - a_i) \left(\frac{\frac{N}{2} - n_{i-1}^{C \uparrow}}{n_i^{C \uparrow} - n_{i-1}^{C \uparrow}} \right) \right]$$

Où

- $(b_i - a_i)$ est l'amplitude de la classe médiane.
- $n_i^{C \uparrow}$ est l'effectif cumulé croissant de la classe médiane
- $n_{i-1}^{C \uparrow}$ est l'effectif cumulé croissant de la classe précédente

Pour l'exemple 1, $\frac{N}{2} = \frac{20}{2} = 10$, $Me \in$ La classe médiane $[30; 40[$

La valeur de la médiane est $Me = 30 + \left[10 \left(\frac{10-7}{13-7} \right) \right] = 35$ Pounds

4– Quartiles

Déterminons le premier quartile Q_1 , en examinant les fréquences absolues cumulées $n_i^{C \uparrow}$, rappelons que l'on doit atteindre 25% ($\frac{N}{4}$) des valeurs les plus basses.

$$Q_1 = a_i + \left[(b_i - a_i) \left(\frac{\frac{N}{4} - n_{i-1}^{C \uparrow}}{n_i^{C \uparrow} - n_{i-1}^{C \uparrow}} \right) \right]$$

Le calcul du troisième quartile Q_3 est du même type, dans laquelle on atteint 75% ($\frac{3N}{4}$) des valeurs les plus basses.

$$Q_3 = a_i + \left[(b_i - a_i) \left(\frac{\frac{3N}{4} - n_{i-1}^{C \uparrow}}{n_i^{C \uparrow} - n_{i-1}^{C \uparrow}} \right) \right]$$

Pour l'exemple 1, on a:

$$\frac{N}{4} = \frac{20}{4} = 5, \text{ d'où}$$

$$Q_1 \in [20; 30[\Rightarrow Q_1 = 20 + \left[10 \left(\frac{5 - 4}{7 - 4} \right) \right] = 23.33 \text{ Pounds}$$

et $\frac{3n}{4} = \frac{3 \times 20}{4} = 15$, d'où

$$Q_3 \in [40; 50[\Rightarrow Q_3 = 40 + \left[10 \left(\frac{15 - 13}{18 - 13} \right) \right] = 44 \text{ Pounds}$$

Paramètres de dispersion (variabilité):

1 – La distance interquartile : $Q_3 - Q_1 = 46.66 - 23.33 = 23.33$.

2 – La variance et l'écart type

Pour La variance et l'écart-type, la remarque qui nous a permis d'utiliser les centres de classes pour le calcul de la moyenne nous servira aussi pour développer un raisonnement analogue au sujet de la variance.

$$\sigma^2 = \text{var}(x) = \frac{1}{N} \sum_i n_i (x_i - \bar{x})^2 = \left(\frac{1}{N} \sum_i n_i (x_i)^2 \right) - (\bar{x})^2$$

et

$$\sigma = \sqrt{\text{var}(x)}$$

Dans l'exemple 01,

$$\text{var}(x) = 159$$

$$\sigma = \sqrt{\text{var}(x)} = \sqrt{159} = 12.6095$$

3 – L'étendue :

$$e = x_{max} - x_{min},$$

4 – Le coefficient de variation :

$$CV = \frac{\text{L'écart type}}{\text{La moyenne}} \cdot 100\% = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Dans l'exemple **01**,

$$\begin{aligned} e &= 60 - 10 = 50 \\ CV &= \frac{12.6095}{34} \cdot 100\% = 37.09\% \end{aligned}$$

Remarque 01 :

Dans le cas des classes d'amplitudes différentes, on doit utiliser les effectifs rectifiés

$$h_i = \frac{n_i}{\alpha_i}$$

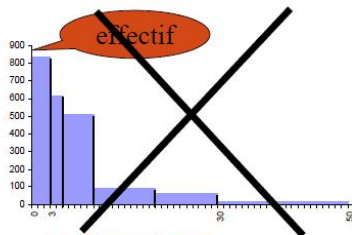
Où $\alpha_i = b_i - a_i$ est l'amplitude de la classe.

L'utilisation des effectifs rectifiés intervient dans 2 choses:

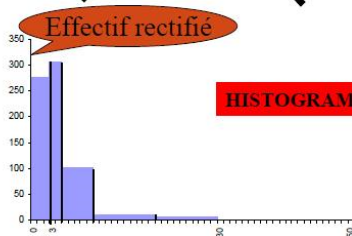
- La Représentation graphique du caractère (**l'histogramme**).
- Le calcul du **mode**.

Exemple 02:

Classes	Effectifs
$[0 - 3[$	830
$[3 - 5[$	615
$[5 - 10[$	510
$[10 - 20 [$	92
$[20 - 30[$	63
$[30 - 50[$	15



Classes	Effectifs n_i	Amplitude a_i	Effectifs rectifiés n_i / a_i
$[0 - 3[$	830	3	276,7
$[3 - 5[$	615	2	307,5
$[5 - 10[$	510	5	102,0
$[10 - 20 [$	92	10	9,2
$[20 - 30[$	63	10	6,3
$[30 - 50[$	15	20	0,75



Le Mode :

Dans ce cas, pour calculer le mode, il faut appliquer la même formule précédente, mais en calculant d_1 et de d_2 à l'aide des effectifs rectifiés h_i .

Pour l'exemple **02**, La classe modale est $[3; 5[$

$$\left\{ \begin{array}{l} d_1 = h_i - h_{i-1} = 307.5 - 276.7 = 30.8 \\ d_2 = h_i - h_{i+1} = 307.5 - 102 = 205.5 \\ a_i = 3 \text{ et } b_i = 5 \\ b_i - a_i = 5 - 3 = 2 \end{array} \right.$$

Le mode est $M_O = 3 + \left[2 \left(\frac{30.8}{30.8+205.5} \right) \right] = 3.26$.

Remarque 02 : Construction des classes pour une série statistique :

- Si les observations sont liées l'une à l'autre ou très nombreuses, on peut les répartir en classes:
 - Le nombre de classes est $C \approx \sqrt{N}$ où N est le nombre total d'observations
 - L'amplitude des classes est

$$\alpha = \frac{e}{C} = \frac{x_{\max} - x_{\min}}{C}$$

Exemple: Voir l'exercice 03 de la série 02 au TD.

Exercice

Sur une population issue d'une certaine région, on a mesuré le taux de cholestérol (en *g/litre*), on a obtenu les résultats suivants :

Classes	Effectif
$[0, 84; 0, 96[$	2
$[0, 96; 1, 08[$	8
$[1, 08; 1, 20[$	16
$[1, 20; 1, 32[$	19
$[1, 32; 1, 44[$	2
$[1, 44; 1, 56[$	2
$[1, 56; 1, 68[$	1

1. Quel est le nombre d'individus de cette population?
2. Quel est le caractère étudié ?
3. Déterminer la moyenne arithmétique, le mode, la médiane et les quartiles Q_1 et Q_3 .
4. Déterminer l'écart-type.
5. Que peut-on conclure?.