PARTIE III) Caractère continu

Définition : un caractère continu est un caractère dont Les modalités sont données sous forme de classes, notées $[e_i , e_{i+1}[$ i= 1, 2, …,k. telles que :

*L'amplitude de la classe est notée $a_i = e_{i+1} – e_i$.

*Le centre de la classe est $c_i = (e_i + e_{i+1}) /2$.

III.1) Représentation graphique : dans ce cas il existe 3 types de représentations graphiques :

a) ***Histogramme :***
Dans le cas d'une variable statistique continue, le diagramme des effectifs (ou fréquences) s'appelle histogramme. A chaque classe est associée un rectangle dont la largeur est l'amplitude de la classe et dont la hauteur est l'effectif (ou la fréquence) de cette classe.

b) ***Polygone des effectifs*** (ou fréquences)
Le polygone est la ligne brisée joignant les milieux des bases supérieures des différents rectangles.

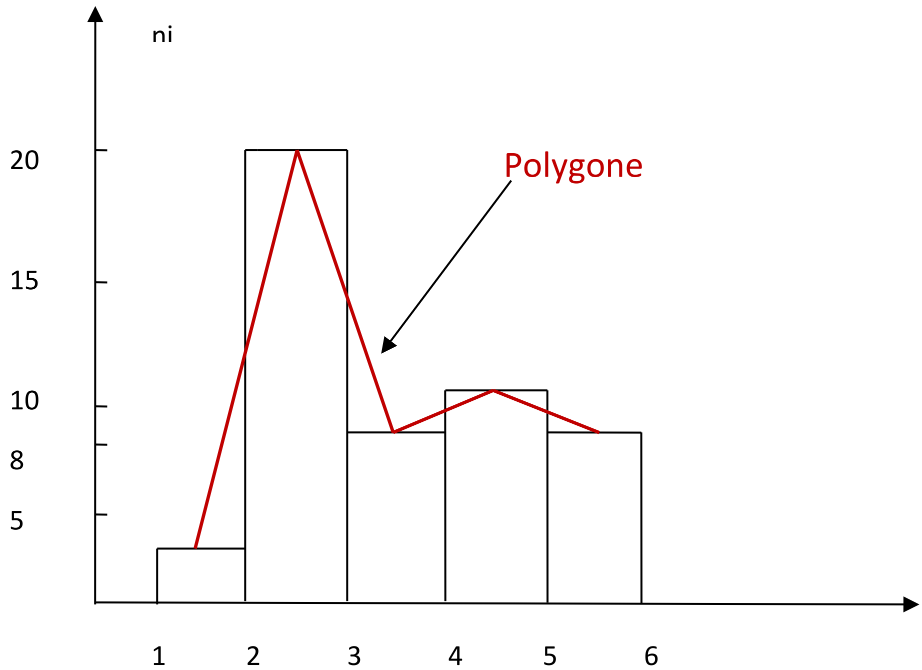
c) **Courbe cumulative** : c'est la représentation graphique des effectifs cumulés Ni ou fréquences cumulées Fi.
Dans ce cas, la Courbe cumulative est une ligne brisée faite de segments de droites joignant les points ayant pour abscisse, la limite supérieure d'une classe et pour ordonnée la fréquence cumulée (ou effectif cumulé)de cette classe.

Exemple A : Une étude pourtant sur la durée de vie de 50 ampoules est résumée dans le tableau suivant :

durée de vie (xi)	Nombre d'ampoules (ni)	Effectifs cumulés (Ni)
[1,2[	4	4
[2,3[	20	24
[3,4[	8	32
[4,5[	10	42
[5,6[	8	50

Donner les 3 représentations graphiques caractérisant cette série.

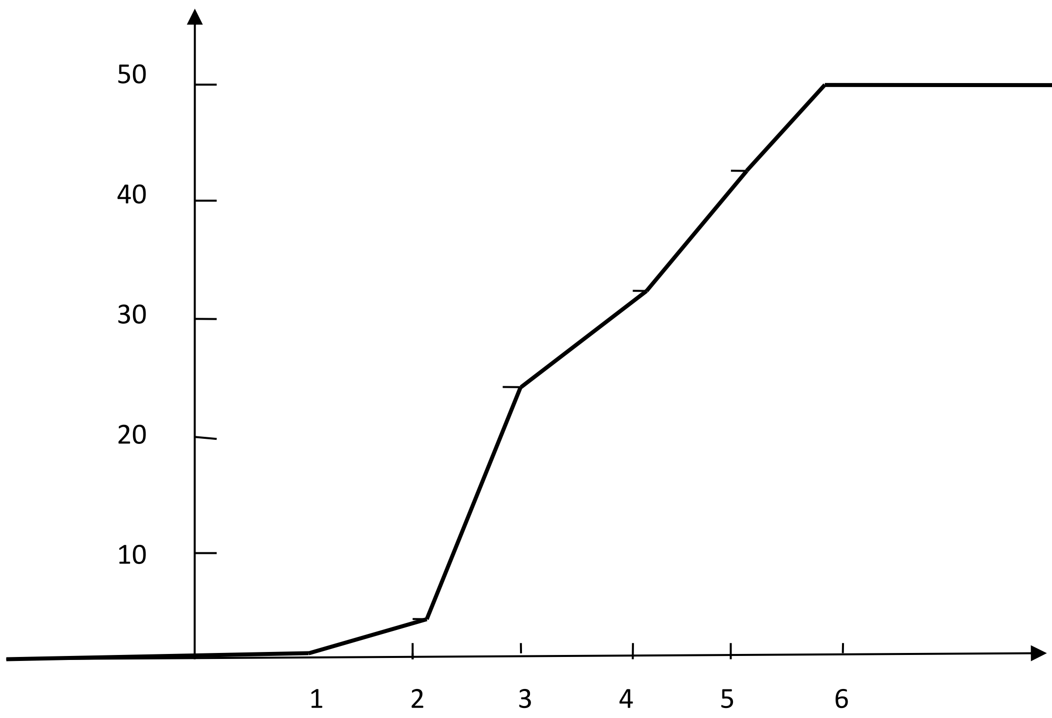
a)Histogramme :



Histogramme de la variable « durée de vie »

b) polygone :

c) la courbe cumulative



III.2) Regroupement en classes :

Souvent, dans une série statistique discrète, plusieurs valeurs sont très proches les unes des autres, au lieu de les étudier séparément, on les regroupe dans des classes d'amplitudes égales, $[e_1,e_2[$, $[e_2,e_3[$ …tel que :

- Le nombre de classes, noté k est l'entier le plus proche de racine n. n étant la taille de la série.
- l'amplitude a doit vérifier : a ≥ e / k .e étant l'étendue de la série.

  Exemple : soit une série de 40 valeurs telles que X_{min} =150 et X_{max} =185

  Regroupement en classes :

  1) e =185 – 150 = 35
  2) nombre de classe k≈ $\sqrt{n}$ =$\sqrt{40}$ = 6.3 donc k = 6
  3) l'amplitude a : a ≥ e / k ; a ≥ 35 / 6 donc a =6

     Les classes sont [150,156[, [156,162[, [162,168[, [168,174[,[174,180[, [180,186[.

III.3) paramètres de position : Sont des paramètres permettant de résumer une série d'observations par une seule valeur.

III.3.1) le Mode (Mo) :c'est la valeur de la variable statistique pour laquelle l'effectif, ou la fréquence, est la plus élevée. On l'appelle aussi valeur dominante et classe modale dans le cas continu.

***Détermination du mode dans le cas d'une variable continue :**

Après localisation de la classe modale, soit Mo ∈ [ei, e $_{i+1}$[; le mode se calcule par la formule suivante :

$$\text{Mo} = ei + \frac{\propto 1}{\propto 1 + \propto 2}\, ai$$

Où :

ei : désigne la borne inférieure de la classe modale

ai : l'amplitude de la classe modal$e$.

$\propto 1$: La différence entre les effectifs de la classe modale et de la classe précédente

$\propto 2$: La différence entre les effectifs de la classe modale et de la classe suivante

Exemple A : Mo ∈ [2, 3[; le mode se calcule par la formule suivante :

$$\text{Mo} = 2 + \frac{16}{16+12}\,.1$$

($\propto 1 = 20 - 4 = 16$ et $\propto 2 = 20 - 8 = 12$)

III.3.2) la médiane (Me): ma médiane, notée Me, est la valeur de la variable statistique qui partage la série d'observations ordonnées en deux groupes d'effectifs égaux.

***Détermination analytique de la médiane dans le cas continu** : dans ce cas on détermine d'abord la classe médiane vérifiant N(Me) = n/2 OU F(Me) = 0.5 , soit Me ∈ [ei, ei+1[; la médiane se calcule par la formule suivante :

$$\text{Me} = ei + \frac{\frac{n}{2} - Ni-1}{ni}\, ai$$

Où :

e_i : désigne la borne inférieure de la classe médiane
a_i : l'amplitude de la classe médiane.
N_{i-1} : l'effectif cumulé avant la classe médiane
n_i : désigne l'effectif de la classe médiane

Exemple : On reprend l'exemple A sur la durée de vie de 50 ampoules :

N(Me) = n/2 = 25 donc Me ∈ [3, 4[,

$$Me = 3 + \frac{25-24}{8} .1 = 3.12$$

<u>*Détermination graphique de la médiane</u>

La médiane est la solution de l'équation N(Me) = n /2 (respectivement F(Me) =0.5) ; Sur la courbe cumulative des effectifs cumulés (respectivement des fréquences cumulées), la médiane est l'abscisse du point d'ordonnée n/2 (respectivement 1/2),

III.3.3) les quartiles (Q1, Q2, Q3): sont les valeurs de la variable statistique qui partagent la série ordonnée en 4 parties de mêmes effectifs, notés Q1, Q2 (Me)et Q3.

<u>*Détermination des quartiles dans le cas continu :</u>

* N(Q1) = n/4, soitQ1 ∈ [e_i, e_{i+1}[; Q1 se calcule par la formule suivante :

$$Q1 = ei + \frac{\frac{n}{4} - Ni-1}{ni}\, ai$$

Où :
e_i : désigne la borne inférieure de la classe de Q1
a_i : l'amplitude de la classe [e_i, e_{i+1}[.
N_{i-1} : l'effectif cumulé avant la classe [e_i, e_{i+1}[
n_i : désigne l'effectif de la classe [e_i, e_{i+1}[

* N(Q3) = 3n/4, soitQ3 ∈ [ci, ci+1[; Q3 se calcule par la formule suivante :

$$Q3 = ci + \frac{3\frac{n}{4} - Ni-1}{ni}\, ai$$

Où :
e_i : désigne la borne inférieure de la classe de Q3
a_i : l'amplitude de la classe [c_i, c_{i+1}[.
N_{i-1} : l'effectif cumulé avant la classe [c_i, c_{i+1}[
n_i : désigne l'effectif de la classe [c_i, c_{i+1}[

Exemple : On reprend l'exemple A sur la durée de vie de 50 ampoules :

N(Q1) = n/4 = 12.5 donc Q1 ∈ [2, 3[,

$$Q1 = 2 + \frac{12.5-4}{20}.1 = 2.425$$

N(Q3) = 3n/4 = 37.5 donc Q3 ∈ [4, 5[,

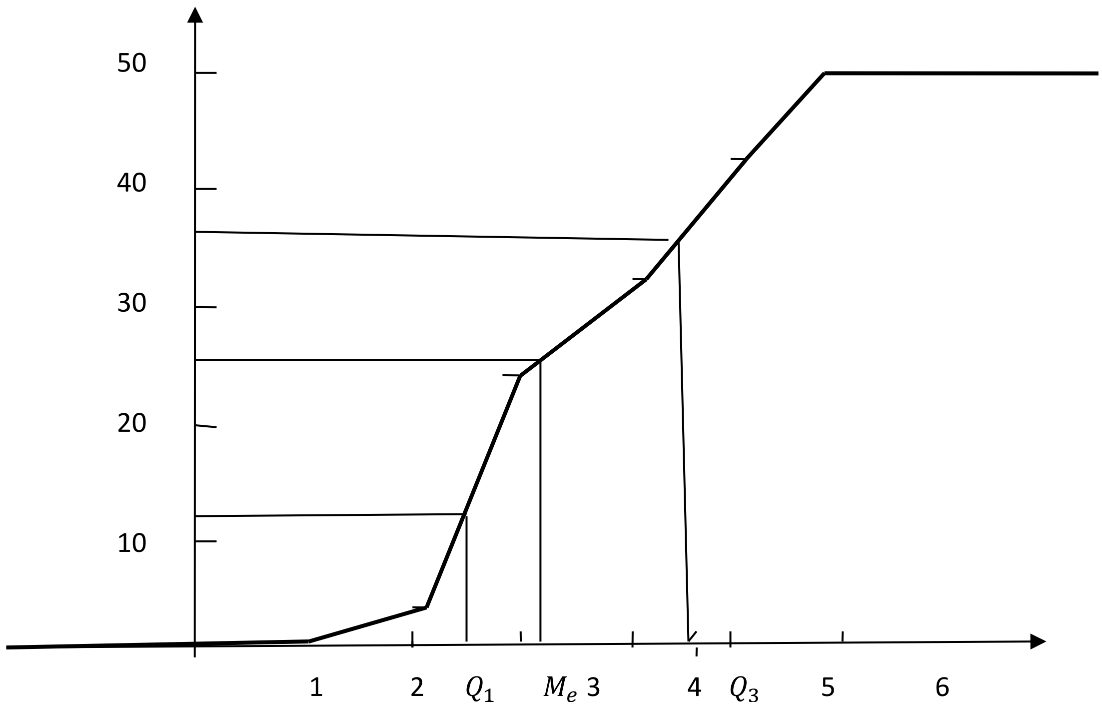
$$Q3 = 4 + \frac{37.5-32}{10}.1 = 4.55$$

***Détermination graphique de Q1 et Q2(Me) et Q3**

-Sur la courbe cumulative des effectifs cumulés (respectivement des fréquences cumulées), Q1 est l'abscisse du point d'ordonnée n/4 (respectivement 1/4),

-Sur la courbe cumulative des effectifs cumulés (respectivement des fréquences cumulées), Q3 est l'abscisse du point d'ordonnée 3n/4 (respectivement 3/4),

Exemple : : On reprend l'exemple A



Définition : On appelle écart interquartile la différence entre les valeurs du troisième et du premier quartile : Q3 – Q1

III.3.4) la moyenne arithmétique $\bar{X}$

Concéderons une série statistique et sa distribution (Xi , ni) ; 1≤ i ≤ k

La moyenne arithmétique de cette série , notée $\bar{X}$ est par définition

$\bar{X} = 1/n \sum_{i=1}^{k} nixi = \sum_{i=1}^{k} fixi$

Remarque : dans le cas continu, on remplace les xi par leur centre Ci, ainsi on obtient : $\bar{X} = 1/n \sum_{i=1}^{k} nici = \sum_{i=1}^{k} fici$

III.4) paramètres de dispersion : ces paramètres permettent de mesurer la dispersion de la série statistique par rapport à ses paramètres de position.

III.4.1) l'étendue (e) : $e = X_{max} - X_{min}$

III.4.2) la variance et l'écart type

Soit une série statistique de n valeurs X_1 , X_2 .....,X_n .

On définit la variance de X (var(X)) par la formule suivante:

$$Var(X) = 1/n \sum_{i=1}^{K} nixi^2 - \bar{X}^2 = \sum_{i=1}^{K} fixi^2 - \bar{X}^2$$

- *l'écart type δx est la racine de la variance :* $\delta x = \sqrt{var(X)}$

Remarque : dans le cas continu, on remplace les classes du caractère par leur centre Ci, ainsi on obtient : $var(X) = \sum_{i=1}^{6} fici^2 - \bar{X}^2$

III.4.3) coefficient de variation

Définition : le coefficient de variation est le rapport de l'écart-type à la moyenne arithmétique $\bar{X}$:

$$Cv = \frac{\delta x}{\bar{X}}$$

Remarque : plus ce coefficient est élevé, plus la dispersion est forte

Exemple: *Une compagnie de taxi s'intéresse au kilométrage effectué par ses véhicules. A cet effet, elle a relevé le kilométrage de 50 de ses taxis pour une matinée de travail*

Classes en km	[10, 20[	[20,30[	[30,40[	[40,60[	[60,90[	[90,130[
Effectifs	*7*	*10*	*20*	*6*	*3*	*4*

1) *Calculer le mode et la médiane.*
2) *Calculer la moyenne , la variance et l'écart-type.*
3) *Calculer* le coefficient de variation. Interpréter le résultat

1)Calcul du mode : déterminant d'abord la classe modale ;Mo € [30,40[.

Mo = ei +$\frac{\propto 1}{\propto 1+\propto 2}$ *ai avec* $\propto 1$*=20-10=10 et* $\propto 2$ *=20-6=14.*

Mo= 30 +$\frac{10}{24}$*. 10 =34.166*

Classes en km	*I*	*Ni*	*Ci*	*ni ci*	*nici*^{2}
[10, 20[	*7*	*7*	*15*	*105*	*1575*
[20,30[	*10*	*17*	*25*	*250*	*6250*
[30,40[	*20*	*37*	*35*	*700*	*24500*
[40,60[	*6*	*43*	*50*	*300*	*15000*
[60,90[	*3*	*46*	*75*	*225*	*16875*
[90,130[	*4*	*50*	*110*	*440*	*48400*
Total	*50*			*2020*	*112600*

**Calcul de la médiane : N(Me) = n/2 = 50/2 =25 donc Me € [30,40[*

$Me = ei + \frac{\frac{n}{2} - Ni-1}{ni}\ ai\ = 30 + \frac{25-17}{20}\ 10 = 34$

**calcul de la moyenne :*$\bar{X}$ *=1/n*$\sum$ *nici =2020/50=40.4*

**Calcul de l'écart type :δx =* $\sqrt{var(X)}$

Var(X)= $1/n \sum_{i=1}^{6} nici^2 - \bar{X}^2$ *=(112600 / 50)- (40.4)² =619.84*

δx =24.8965

3) Cv = $\frac{\delta x}{\bar{X}}$ = $\frac{24.8965}{40.4}$ = 0.61

Interprétation : la dispersion est forte.