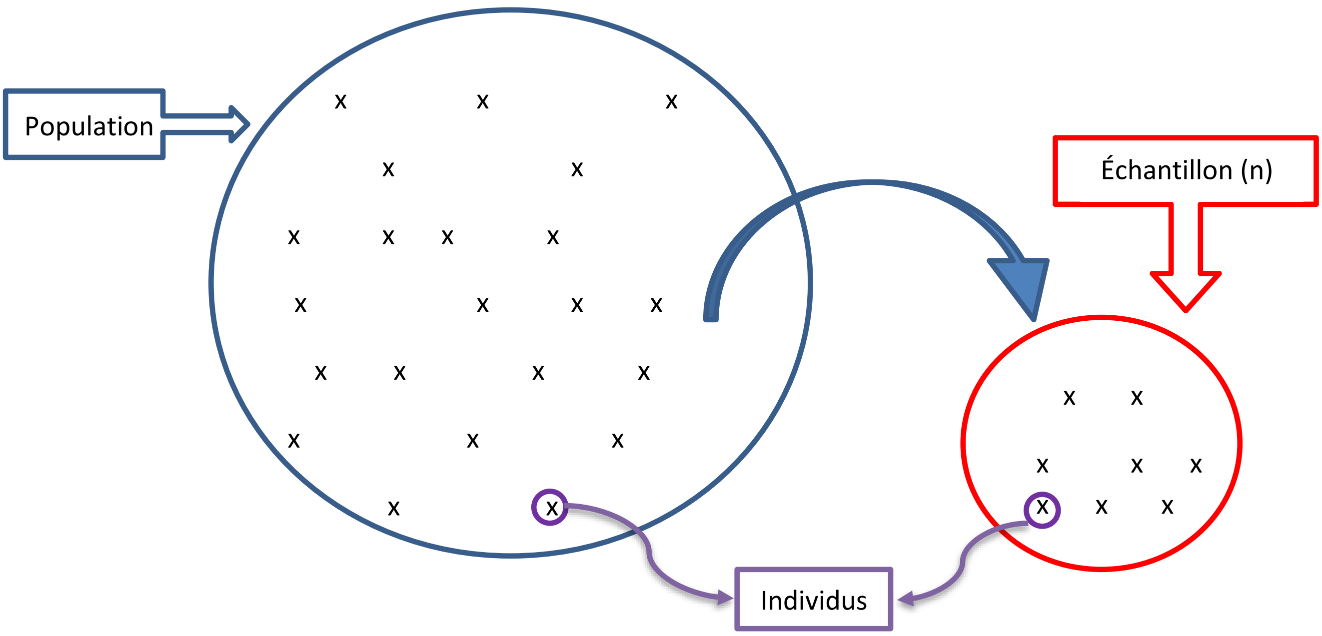
STATISTIQUE DESCRIPTIVE SIMPLE

I / Définitions :

On veut étudier les données relatives aux caractéristiques d'un ensemble d'**individus** ou d'objets appelé **population**. Il est difficile d'observer toutes les données lorsque le nombre d'individus de l'ensemble de la population est élevé. On examine un nombre restreint de la population qu'on appelle **échantillon**. Chaque individu peut être étudié relativement à un ou plusieurs **caractères**. Un caractère est appelé **variable statistique** qu'on note par ***X***. L'ensemble des données observées relatives à un caractère constitue une **série statistique**. Un caractère peut présenter plusieurs **modalités**. On suppose qu'il existe $\boldsymbol{k}$ modalités notées $\{X_1, X_2, \dots, X_k\}$.



1. Caractère qualitatif : Si ses modalités sont des catégories et ne sont pas mesurables. Il existe deux types de caractères qualitatifs :

- Ordinal s'il peut être ordonné, comme degré de brulure (1er, 2ème ou 3ème degré) ou niveau de formation.

- Nominal s'il ne s'ordonne pas, comme profession ou couleur des yeux.

2. Caractère quantitatif : Si ses modalités sont mesurables, c'est-à-dire, on peut correspondre un nombre à chaque modalité. On distingue deux types :

a. Caractère quantitatif discret : S'il ne peut prendre que des valeurs isolées d'un certain intervalle, par exemple le nombre d'enfants d'une famille $\in \{0,1, \dots, n\}$.

b. Caractère quantitatif continu : S'il peut prendre toute valeur appartenant à un intervalle (les valeurs possibles du caractère sont des nombres réels), par exemple la taille ou le poids des nouveau-nés $\in \mathbb{R}$.

II/ Séries statistiques :

On appelle **série statistique** la suite des valeurs prises (mesurées ou observées) par une variable ***X*** sur les individus. Le nombre d'individus est noté ***n***, et les valeurs observées de la variable ***X*** sont notées $\{x_1, x_2, \dots, x_n\}$.

Exemples :

1/ On s'intéresse à la variable **état-civil** notée ***X*** et à la série statistique des valeurs prises pour 10 personnes. Les modalités sont : (**C** : célibataire, **M** : marié, **V** : veuf, **D** : divorcé). Soit la série statistique observée :

M M V C C M C C M C. ici $n = 10$ et $x_1 = M, x_2 = M, x_3 = V, \dots, x_{10} = C$.

2/ Le caractère ***X*** : le nombre d'enfants par famille. La série statistique :

3 2 0 1 1 3 3 2 5 0 2 2 1 3 4. $x_1 = 3, x_2 = 2, x_3 = 0, \dots, x_{15} = 4.\ n = 15$.

1/ Effectif, fréquence effectif cumulé et fréquence cumulée :

On appelle **effectif** d'une modalité, le nombre de fois que cette modalité apparaît dans la série statistique. On note $\boldsymbol{n_i}$ l'effectif de la modalité X_i, $i = 1, \dots, k$.

La **fréquence** d'une modalité est la proportion correspondante $\boldsymbol{f_i} = \frac{n_i}{n}$, $i = 1, \dots, k$.

On note $\boldsymbol{n_{ic}}$ **l'effectif cumulé** et $\boldsymbol{f_{ic}}$ la **fréquence cumulée** de la modalité X_i.

$\boldsymbol{n_{ic}} = n_1 + n_2 + \cdots + n_i$ et $\boldsymbol{f_{ic}} = f_1 + f_2 + \cdots + f_i$.

<u>2/ Tableau statistique :</u> Dans un tableau statistique on résume les données en portant les modalités et les fréquences correspondantes.
<u>Exemples</u> Dans le cas qualitatif, on reprend l'exemple du caractère état civil :

X_i : état civil	n_i	f_i
C	5	0.5
M	4	0.4
V	1	0.1
D	0	0
Totaux	10	1

M M V C C M C C M C

Dans le cas quantitatif discret, on reprend l'exemple du nombre d'enfants par famille.

Nbre d'enf	n_i	n_{ic}	f_i	f_{ic}
0	2	2	2/15	2/15
1	3	5	3/15	5/15
2	4	9	4/15	9/15
3	4	13	4/15	13/15
4	1	14	1/15	14/15
5	1	15	1/15	1
Totaux	15	/	1	/

3 2 0 1 1 3 3 2 5 0 2 2 1 3 4

<u>- Cas continu :</u> Puisque les modalités sont infinies alors on fait une répartition des données dans des intervalles appelés classes et on note par X_i le centre de la i ème classe. Si le nombre de classes k n'est pas donné, on prend $k = E[\sqrt{n}]$.

Soit $[l_{i-1}, l_i[$ la i ème classe, $i = 1, \dots, k$. On note $e = x_{max} - x_{min}$ étendue de la série. Alors l'amplitude de cette classe est : $a = l_i - l_{i-1} = \frac{e}{k}$ et $l_0 = x_{min}$

<u>Remarque :</u> Choisir l'amplitude de sorte que toutes les valeurs de la série soient incluses dans le tableau.

<u>Exemple :</u> On mesure la taille en cm de 50 élèves d'une classe.

152	152	152	153	153
154	154	154	155	155
156	156	156	156	156
157	157	157	158	158
159	159	160	160	160
161	160	160	161	162
162	162	163	164	164
164	164	165	166	167
168	168	168	169	169
170	171	171	171	171

$e = x_{max} - x_{min} = 171 - 152 = 19$, $k = E\left[\sqrt{n}\right] = 7.$

$a = \frac{e}{k} = 2.7$ on prend $a = 3$.

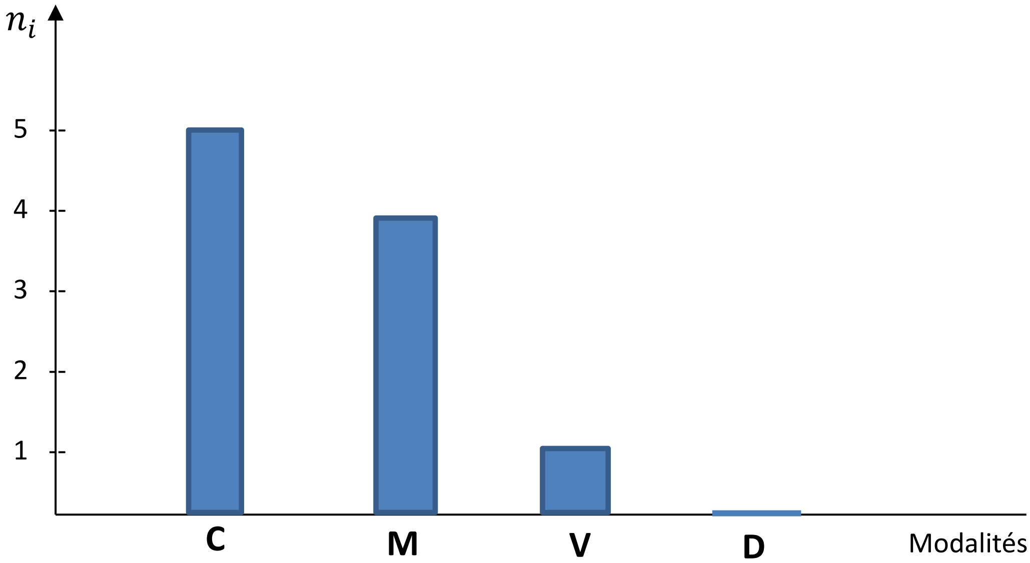
Classes	X_i	n_i	n_{ic}	f_i	f_{ic}
[152, 155[	153.5	8	8		
[155, 158[	156.5	10	18		
[158, 161[	159.5	9	27		
[161, 164[	162.5	6	33		
[164, 167[	165.5	6	39		
[167, 170[	168.5	6	45		
[170, 173[	171.5	5	50		
Totaux	/	50	/		

III/ Représentation graphique : Les données statistiques se présentent avant tout traitement sous forme désordonnée. Après une mise en ordre et la présentation de ces données sous forme de tableau, les représentations graphiques permettent de **visualiser** la distribution.

1/ Variable qualitative : Le principe consiste à représenter les données par des diagrammes dont les différentes parties ont des aires proportionnelles aux effectifs ou aux fréquences. (On ne représente pas les n_{ic} et les f_{ic}).

- **Diagramme à bandes ou tuyaux d'orgue :**

Dans le cas qualitatif, les modalités sont portées sur une droite horizontale qui n'est pas orientée car elles ne sont pas mesurables. Les effectifs sont sur un axe vertical et la hauteur de chaque bande est proportionnelle à n_i ou f_i.

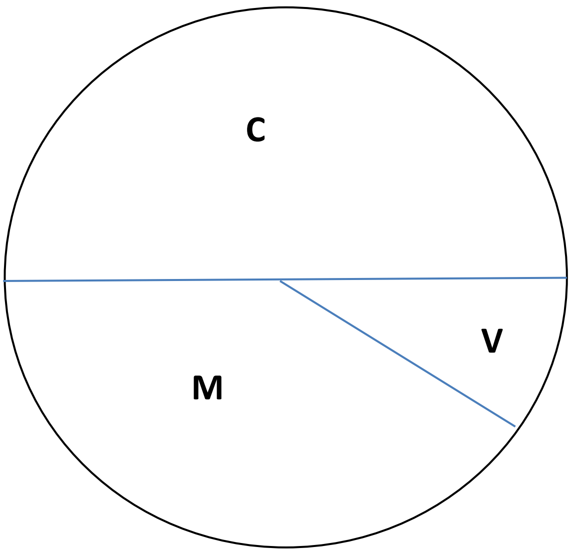


- **Diagramme en secteurs circulaires :** (Camembert)

C'est le plus utilisé. On représente chaque modalité dans un secteur circulaire dont l'angle $\alpha_i = f_i \times 360°$.

Exemple :

Etat civil	n_i	f_i	α_i
C	5	0.5	180°
M	4	0.4	144°
V	1	0.1	36°
D	0	0	0°
Totaux	10	1	360°



2/ variable quantitative :

a/ cas discret :

- **Diagramme en bâtons :** On porte en abscisse les différentes modalités du caractère X et en ordonnée les effectifs n_i (ou les fréquences f_i).

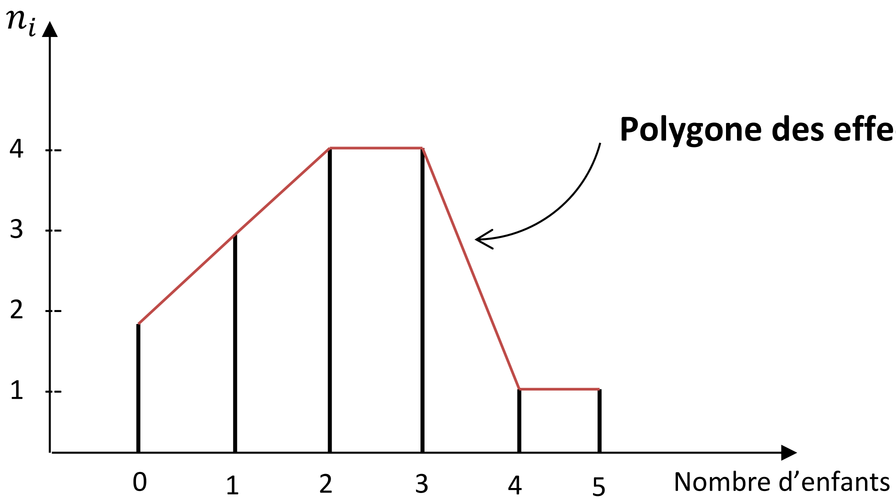


Diagramme en bâtons

- **Polygone des effectifs :** (ou des fréquences) C'est la ligne brisée qui joint les sommets des bâtons.
- **Diagramme cumulatif :** On porte en abscisse les valeurs X_i du caractère X et en ordonnée les effectifs (ou fréquences) des individus pour lesquels le caractère X est inférieur ou égal à X_i. On obtient une courbe en escalier.

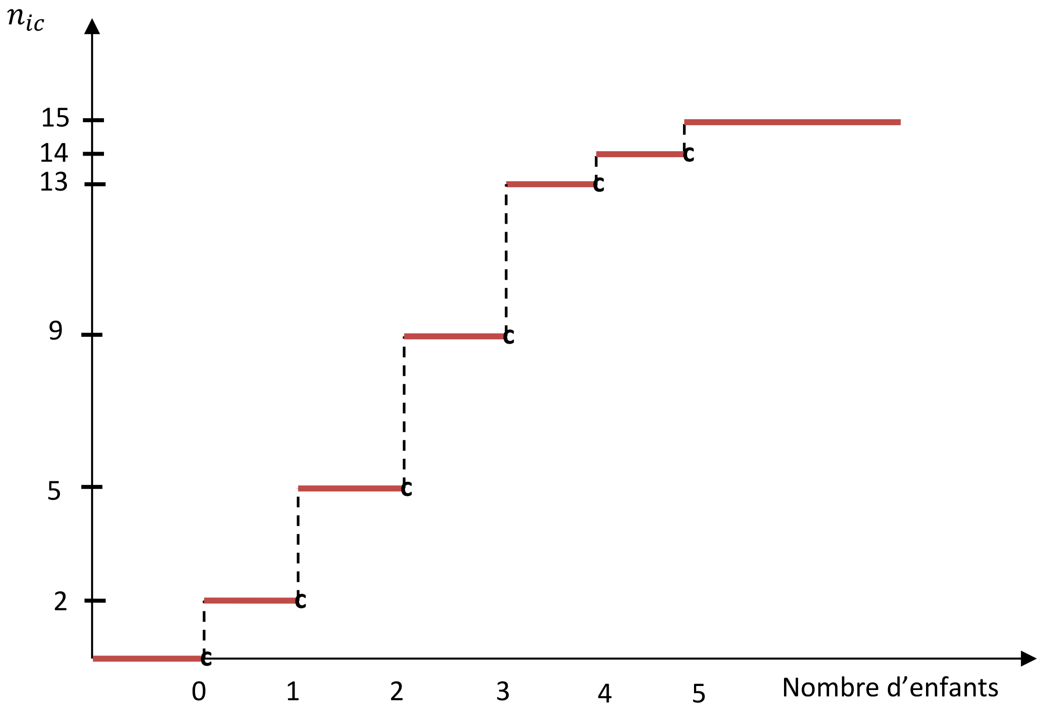
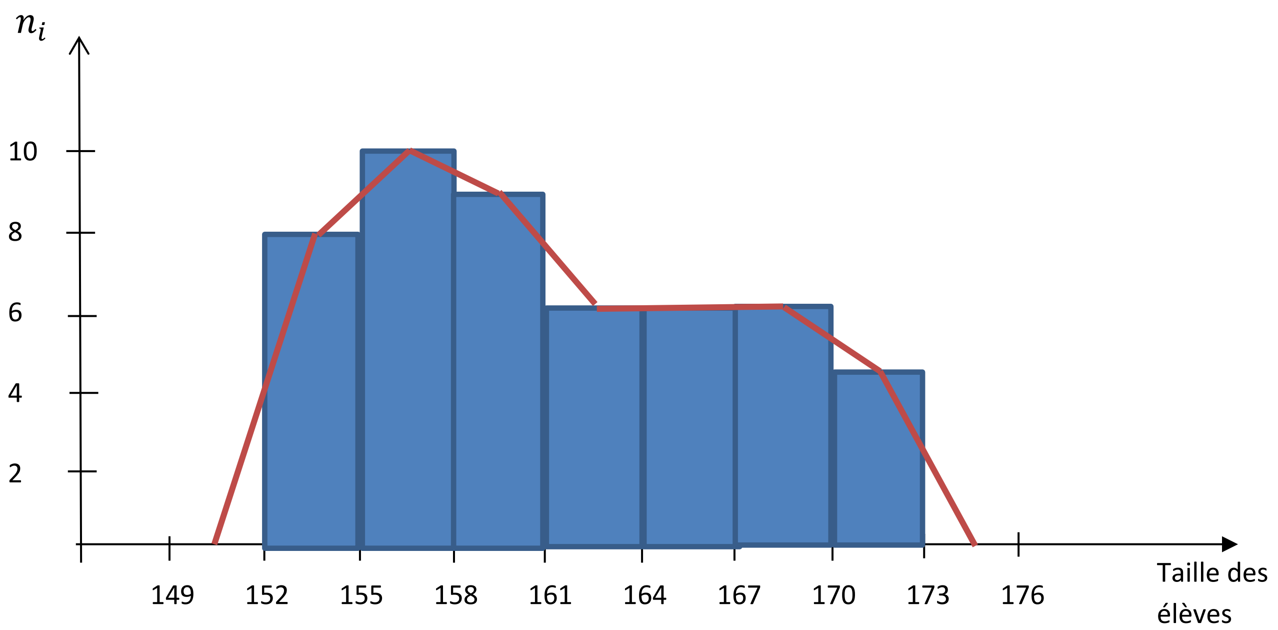


Diagramme cumulatif

b/ Cas continu :

- **Histogramme :** A chaque classe on associe un rectangle dont la largeur est l'amplitude de la classe et dont la hauteur est l'effectif (ou fréquence) qui correspond.

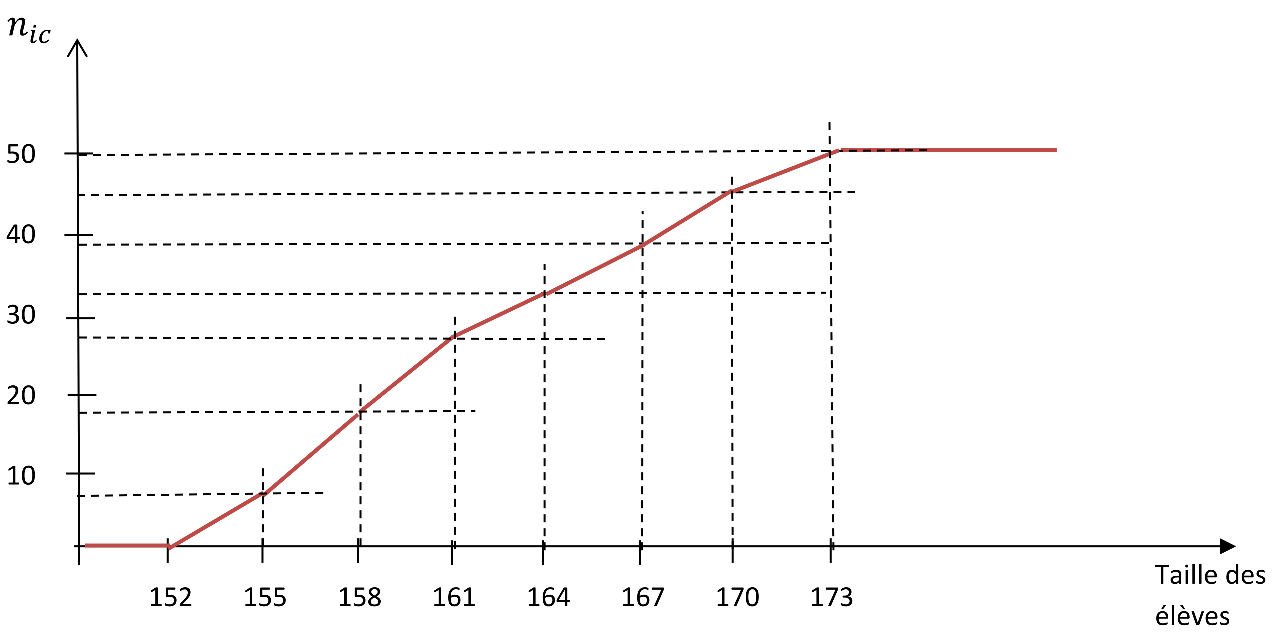


Histogramme

Classes	X_i	n_i	n_{ic}	f_i	f_{ic}
[152, 155[	153.5	8	8		
[155, 158[	156.5	10	18		
[158, 161[	159.5	9	27		
[161, 164[	162.5	6	33		
[164, 167[	165.5	6	39		
[167, 170[	168.5	6	45		
[170, 173[	171.5	5	50		
Totaux	/	50	/		

- **Polygone des effectifs :** c'est la ligne brisée qui passe par les points milieux des sommets des rectangles. Deux classes fictives ont été rajoutées $[149,152[$ et $[173,176[$, dont les effectifs sont nuls, pour que le polygone rejoigne l'axe des abscisses.

- **Polygone des effectifs cumulés :** (ou des fréquences cumulées). C'est la ligne brisée joignant les points obtenus en portant en ordonnée au côté droit de chaque classe (limite supérieure en abscisse), l'effectif cumulé (ou fréquence cumulée).



Polygone des effectifs cumulés

Classes	X_i	n_i	n_{ic}	f_i	f_{ic}
[152, 155[	153.5	8	8	8/50	8/50
[155, 158[	156.5	10	18	10/50	18/50
[158, 161[	159.5	9	27	9/50	27/50
[161, 164[	162.5	6	33	6/50	33/50
[164, 167[	165.5	6	39	6/50	39/50
[167, 170[	168.5	6	45	6/50	45/50
[170, 173[	171.5	5	50	5/50	50/50
Totaux	/	50	/		

- **Fonction de répartition :** On note par $F(x)$ la fonction de répartition associée à une valeur x, qui est une application de $\mathbb{R}$ dans $[0,1]$.
- **Cas discret** : C'est la fréquence cumulée des observations inférieures à x.

$$F(x) = \begin{cases} 0 & si \quad x < X_1 \\ f_{ic} & si \quad X_i \leq x < X_{i+1} \text{ pour } i = 1, \dots, k. \\ 1 & si \quad x \geq X_k \end{cases}$$

- **Cas continu :**

$$F(x) = \begin{cases} 0 & si \quad x < l_0 \\ \sum_{j=1}^{i-1} f_j + \frac{x - l_{i-1}}{l_i - l_{i-1}} f_i & si \quad l_{i-1} \leq x < l_i \text{ pour } i = 1, \dots, k. \\ 1 & si \quad x \geq l_k \end{cases}$$

Exemples :

- Taille de 50 élèves. Calculer la fonction de répartition pour $x = 160$.

$$F(160) = \frac{18}{50} + \frac{160 - 158}{161 - 158} \times \frac{9}{50} = \frac{24}{50} = 0.48$$

- Nombre d'enfants de 15 familles. Calculer la fonction de répartition pour $x = 2.3$

Nbre d'enf	n_i	n_{ic}	f_i	f_{ic}
0	2	2	2/15	2/15
1	3	5	3/15	5/15
2	4	9	4/15	9/15
3	4	13	4/15	13/15
4	1	14	1/15	14/15
5	1	15	1/15	1

$$F(2.3) = \frac{9}{15}$$

IV/ paramètres de position :

Ces paramètres sont des valeurs caractéristiques qui permettent une représentation condensée de l'information contenue dans la série statistique.

1/ Mode (Mo) : C'est la valeur que prend le caractère étudié ayant le plus grand effectif. Une série statistique peut être unimodale ou plurimodale.

Cas discret : Le mode est **la modalité** du caractère qui correspond au plus grand effectif. **Exemple** : (nombre d'enfants).

$M_{o1} = 2$ et $M_{o2} = 3$

Cas continu : Il s'agit de classe modale. C'est la classe d'effectif maximum, et elle correspond au sommet de l'histogramme. On peut avoir plusieurs classes modales. Toutes les valeurs de la classe peuvent se réaliser, on détermine une seule valeur qui représentera le mode. On peut choisir par simplicité le centre de la classe modale, cependant il est préférable de faire une interpolation pour tenir compte des classes adjacentes.

n_i : Effectif de la classe modale.

n_{i-1}: Effectif de la classe précédente.

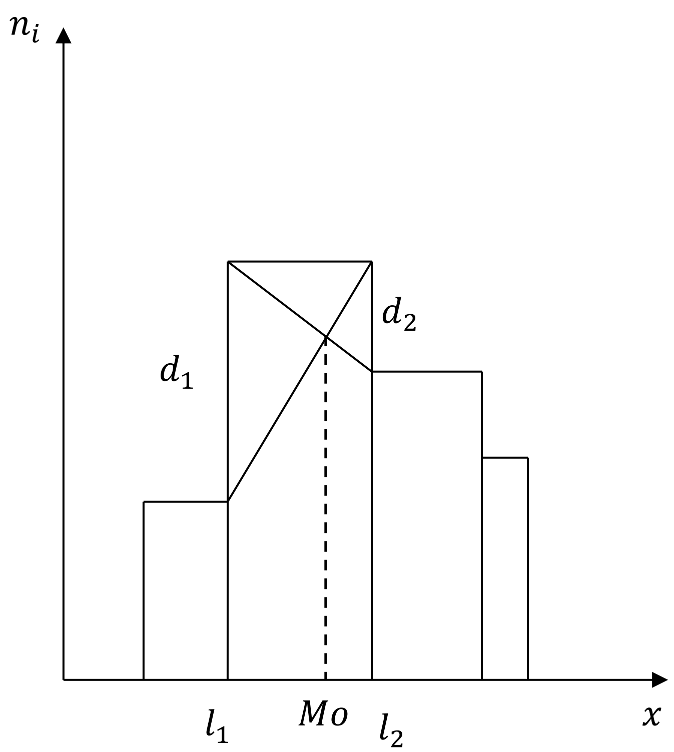
n_{i+1}: Effectif de la classe suivante.

$a = l_2 - l_1$: Amplitude de la classe.

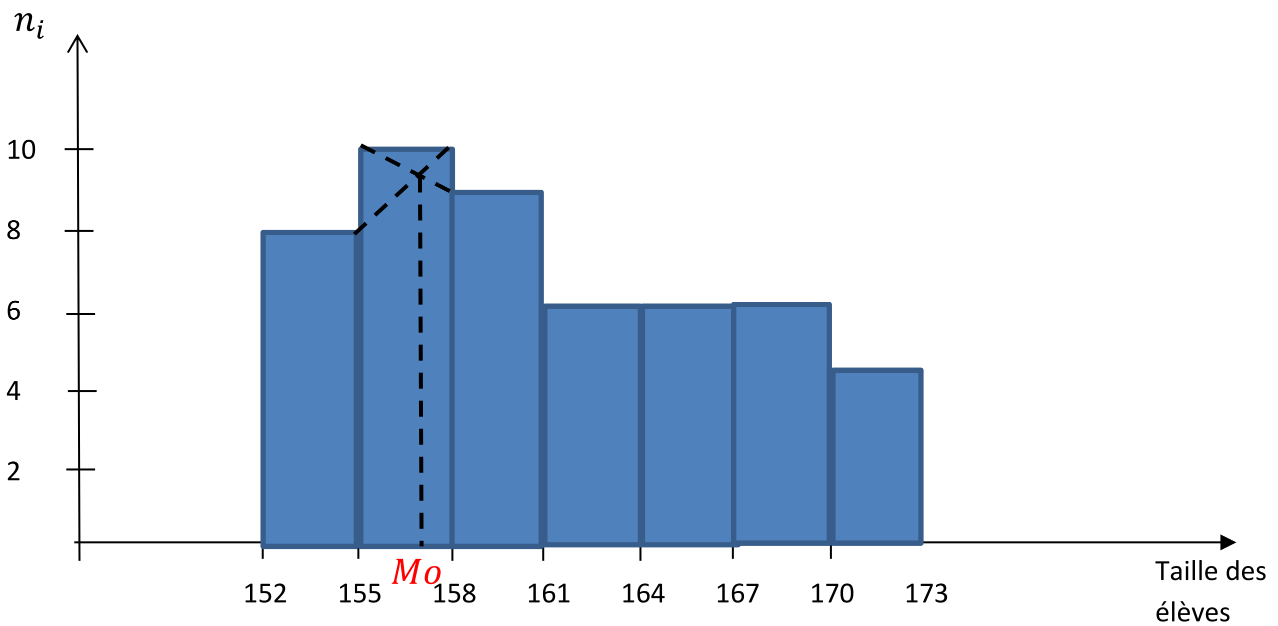
$d_1 = n_i - n_{i-1}$ et $d_2 = n_i - n_{i+1}$.

Alors le mode est :

$$Mo = l_1 + a\frac{d_1}{d_1+d_2}$$



Exemple : (taille de 50 élèves).



Histogramme

$$Mo = 155 + 3 \times \frac{(10-8)}{(10-8)+(10-9)} = 155 + 3 \times \frac{2}{2+1} = 157$$

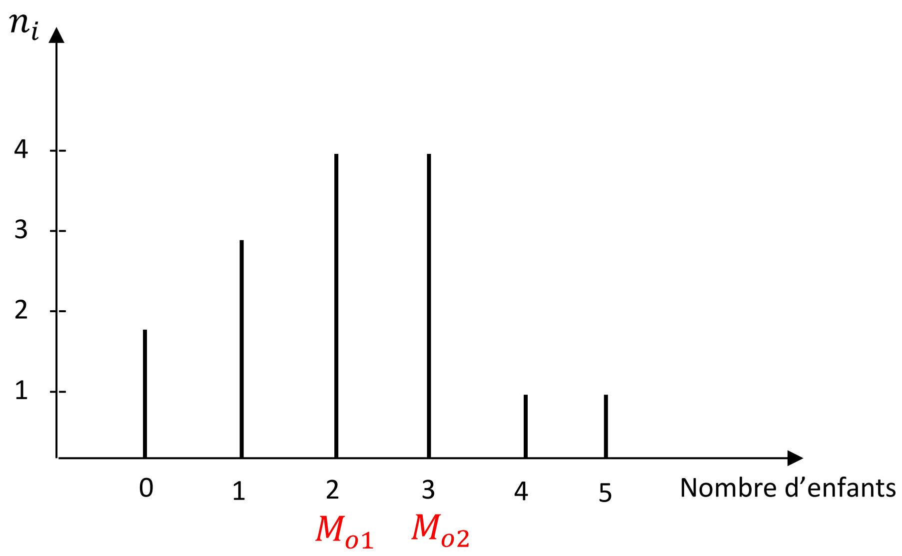
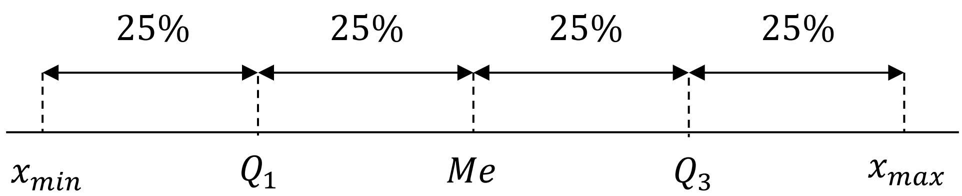


Diagramme en bâtons

$M_{o1} = 2$ et $M_{o2} = 3$

<u>2/ Les quantiles d'ordre α</u> (q_α) :

Soit $\alpha \in [0,1[$. On appelle quantile d'ordre α, noté q_α, la valeur telle qu'il y ait $n\alpha$ (ou α 100 %) observations qui lui soient inférieures dans une série ordonnée de taille *n*. Pour $\alpha = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ on obtient respectivement le 1[er] , 2[ème] , 3[ème] **quartile**, notés $\boldsymbol{Q_1, Q_2, Q_3}$. Le deuxième quartile est appelé **médiane ($\boldsymbol{Me}$).**



3 2 0 1 1 3 3 2 5 0 2 2 1 3 4 $\longrightarrow$ 0 0 1 1 1 2 2 2 2 3 3 3 3 4 5

<u>Cas discret</u> :

$$q_\alpha = \begin{cases} \frac{x_{n\alpha}+x_{n\alpha+1}}{2} & si\ n\alpha \in \mathbb{N} \\ x_{[n\alpha+1]} & si\ n\alpha \notin \mathbb{N} \end{cases}$$

$x_{n\alpha}$ est la valeur x_i qui correspond à $n_{ic} \geq n\alpha$. [.] désigne la partie entière.

On peut déterminer q_α à partir du tableau statistique. Il correspond à la modalité pour laquelle l'effectif cumulé est égal à $n\alpha$ (ou fréquence cumulée égale α).

Exemple :

Pour $\alpha = 0.5$ on a $n\alpha = 7.5 \notin \mathbb{N}$ donc $q_\alpha = x_8 = 2$

Pour $\alpha = 1/3$ on a $n\alpha = 5 \in \mathbb{N}$ donc $q_\alpha = \frac{x_5+x_6}{2} = \frac{1+2}{2} = 1.5$

Nbre d'enf	n_i	n_{ic}	f_i	f_{ic}
0	2	2	2/15	2/15
1	3	5	3/15	5/15
2	4	9	4/15	9/15
3	4	13	4/15	13/15
4	1	14	1/15	14/15
5	1	15	1/15	1

<u>Représentation graphique :</u> Sur le graphe des effectifs (fréquences) cumulés, q_α est l'abscisse du point d'ordonnée $n\alpha$ (ou α) si on a utilisé les n_{ic} (ou les f_{ic}) pour tracer le graphe. Dans le cas discret on a :

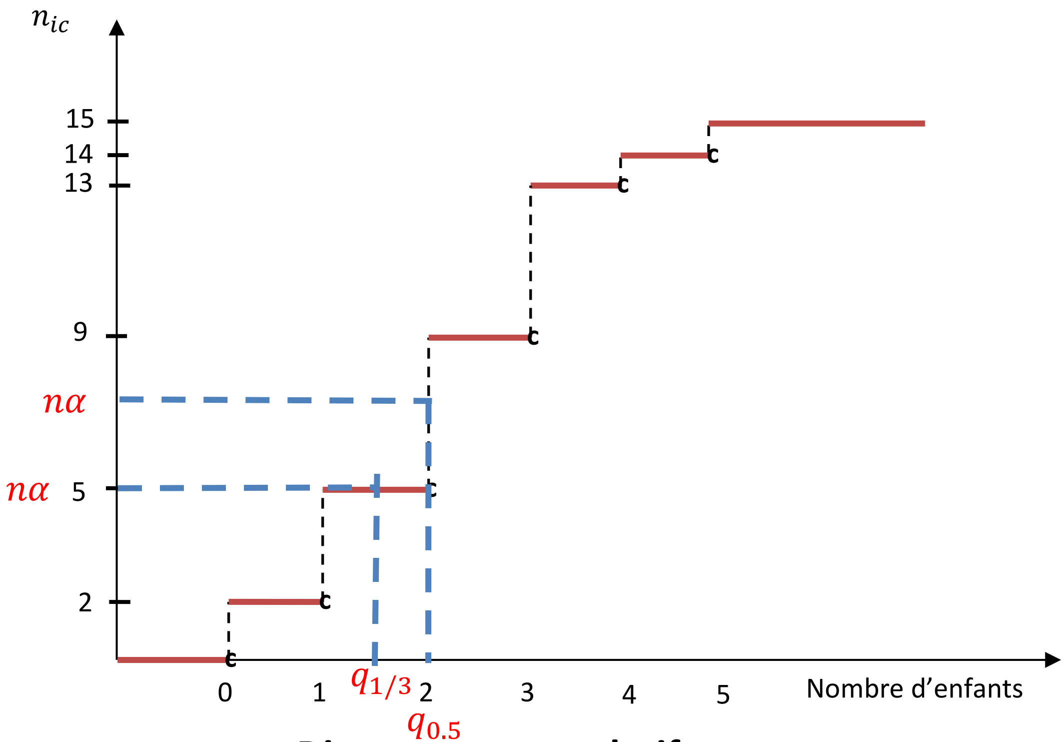


Diagramme cumulatif

<u>Cas continu</u> :

On parlera de classe contenant q_α . C'est la classe telle que $n_{ic} \geq n\alpha$ (ou $f_{ic} \geq \alpha$). q_α est déterminé par interpolation. Soient :

$[a, b[$: classe contenant q_α .

F_2 l'effectif cumulé de $[a, b[$ et F_1 celui de la classe précédente.

$F(b)$ fréquence cumulée de $[a, b[$ et $F(a)$ celle de la classe précédente.

Alors : $q_\alpha = a + (b - a)\frac{n\alpha - F_1}{F_2 - F_1}$ ou $q_\alpha = a + (b - a)\frac{\alpha - F(a)}{F(b) - F(a)}$

__**Exemple :** (taille de 50 élèves).__

- Pour $\alpha = 0.5$ on a $n\alpha = 25$ alors $q_\alpha \in]158, 161]$

$F_1 = 18$ et $F_2 = 27$

$$q_\alpha = a + (b-a)\frac{n\alpha - F_1}{F_2 - F_1} = 158 + 3\,\frac{25-18}{27-18} = 160.33$$

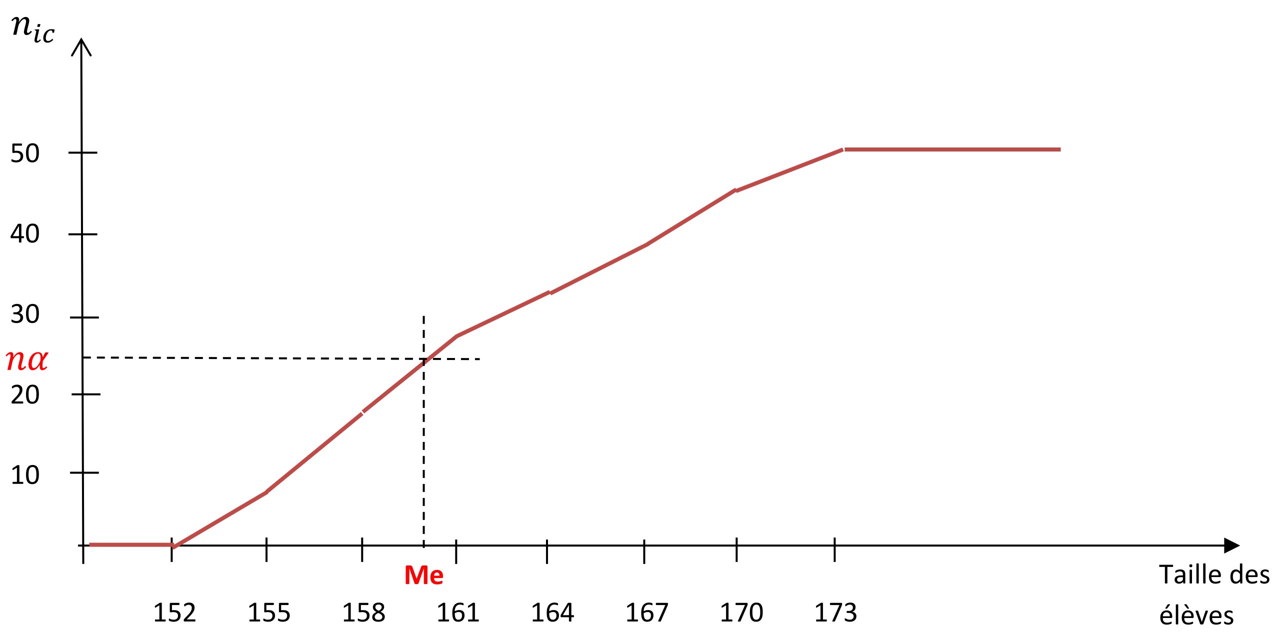
- Pour $\alpha = 1/3$ on a $n\alpha = 16.67$ alors $q_\alpha \in]155, 158]$

$F_1 = 8$ et $F_2 = 18$

$$q_\alpha = a + (b-a)\frac{n\alpha - F_1}{F_2 - F_1} = 155 + 3\,\frac{16.67-8}{18-8} = 157.6$$

Classes	X_i	n_i	n_{ic}	f_i	f_{ic}
[152, 155[	153.5	8	8		
[155, 158[	156.5	10	18		
[158, 161[	159.5	9	27		
[161, 164[	162.5	6	33		
[164, 167[	165.5	6	39		
[167, 170[	168.5	6	45		
[170, 173[	171.5	5	50		
Totaux	/	50	/		

Dans le cas continu on a :



Polygone des effectifs cumulés

3/ La moyenne :

Moyenne arithmétique ($\overline{x}$) : C'est généralement la caractéristique qui représente le mieux **le centre** de la distribution de la série statistique. C'est la caractéristique de **tendance centrale** la plus utilisée. Si les x_i sont les modalités d'une variable discrète ou les centres de classes d'une variable continue, la moyenne arithmétique est :
$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i$ ou bien $\bar{x} = \sum_{i=1}^{k} f_i x_i$
Si $\forall\ i = 1, \dots, k,\ f_i = \frac{1}{n}$ alors on a une moyenne arithmétique simple et si les n_i sont différents alors on a une moyenne arithmétique pondérée.
Les calculs se résument dans le tableau statistique.

Classes	x_i	n_i	n_{ic}	$n_i x_i$
[152, 155[	153.5	8	8	1228
[155, 158[	156.5	10	18	1565
[158, 161[	159.5	9	27	1435.5
[161, 164[	162.5	6	33	975
[164, 167[	165.5	6	39	993
[167, 170[	168.5	6	45	1011
[170, 173[	171.5	5	50	857.5
Totaux	/	50	/	8065

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i = \frac{8065}{50} = 161.3$$

V/ Paramètres de dispersion :

Résumer une distribution statistique par une seule caractéristique telle que le mode, la moyenne ou la médiane, qui représentent une valeur centrale, est insuffisant. La statistique commence là où il y a **variabilité** ; il faut donc définir au minimum une valeur centrale et une **mesure de dispersion** autour de cette valeur centrale.

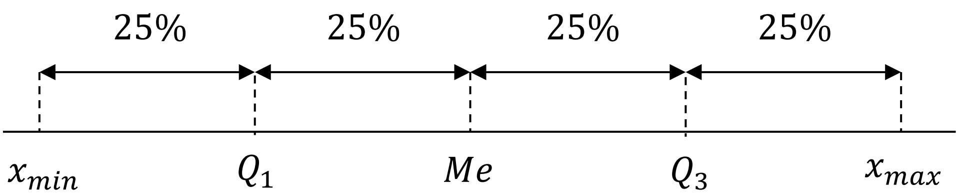
1/ L'étendue (e) : C'est la différence entre la plus grande observation notée x_{max} et la plus petite notée x_{min} . Donc l'étendue est : $e = x_{max} - x_{min}$

$e = x_{max} - x_{min} = 171 - 152 = 19$ pour les données brutes.

$e = x_{max} - x_{min} = 173 - 152 = 21$ pour les données regroupées.

2/ Ecart inter quartile (décile, centile) :

- L'intervalle entre le troisième et le premier quartile $[Q_1, Q_3]$, ou la différence entre eux $Q_3 - Q_1$ est un indicateur de dispersion autour de la médiane Me .



En effet cet indicateur $Q_3 - Q_1$ correspond à un intervalle qui regroupe 50% des observations autour de la médiane.

- L'écart entre le neuvième et le premier décile $D_9 - D_1$ correspond à un intervalle qui regroupe 80% des observations autour de la médiane.
- L'écart entre le quatre-vingt-dix-neuvième et le premier centile contient 98% des observations autour de la médiane.

3/ Les moments :

Ce sont des quantités algébriques qui permettent de décrire les caractéristiques des distributions statistiques : forme, symétrie, aplatissement, tendance centrale, dispersion. La moyenne arithmétique et la variance sont les moments les plus utilisés.

a/ Moment d'ordre r :

Soient x_i , $i = 1, \dots, k$ les modalités d'une variable statistique X discrète ou les centres des classes dans le cas continu. Le moment d'ordre r noté m_r de la variable X est défini par :

$m_r = \frac{1}{n}\sum_{i=1}^{k} n_i\, x_i^r$ (Si $r = 1,\ m_1 = \bar{x}$.)

b/ Moment centré d'ordre r :

Le moment centré d'ordre r noté μ_r par rapport à $\bar{x}$ est défini par :

$\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i\, (x_i - \bar{x})^r$

Si $r = 1,\ \mu_1 = 0.$

Variance et écart-type :

C'est la caractéristique la plus utilisée pour mesurer la dispersion ou l'étalement des données autour de la moyenne. Elle est notée par $V(X)$ ou σ^2 et c'est le moment centré d'ordre 2, μ_2.

$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\,(x_i - \bar{x})^2$ ou $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Remarque : On montre que $\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$

Sur le tableau statistique, on rajoute une colonne pour calculer la variance :

Si a est une constante alors on a les deux propriétés suivantes :

- $V(X + a) = V(X)$
- $V(aX) = a^2 V(X)$

L'écart-type est la racine carrée de la variance et on le note par σ.

Nbre d'enf	n_i	n_{ic}	$n_i x_i$	$n_i \boldsymbol{x_i^2}$
0	2	2	0	0
1	3	5	3	3
2	4	9	8	16
3	4	13	12	36
4	1	14	4	16
5	1	15	5	25
	15	/	32	96

$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{32}{15} = 2.13$ et on a $\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \frac{96}{15} - (2.13)^2 = 1.86$

Classes	x_i	n_i	n_{ic}	$n_i x_i$	$n_i \boldsymbol{x_i^2}$
[152, 155[	153.5	8	8	1228	188498
[155, 158[	156.5	10	18	1565	244922.5
[158, 161[	159.5	9	27	1435.5	-
[161, 164[	162.5	6	33	975	-
[164, 167[	165.5	6	39	993	-
[167, 170[	168.5	6	45	1011	-
[170, 173[	171.5	5	50	857.5	-
Totaux	/	50	/	8065	??????

Changement de variable :

Le calcul de la moyenne et la variance est parfois laborieux à cause des valeurs élevées de X donc de leurs carrées, mais la situation devient particulièrement différente si on effectue le changement de variable suivant :

Si $X = aY + b$ alors $\boldsymbol{\bar{x} = a\bar{y} + b}$ et $\boldsymbol{\sigma_X^2 = a^2\sigma_Y^2}$.

Où $a =$ amplitude de classe et $b =$ milieu de la classe centrale (si le nombre de classes est pair, 2 classes centrales, choisir celle qui a le plus grand effectif).

Sur le tableau statistique, on calcule les valeurs Y_i , n_iY_i , $n_i{Y_i}^2$ puis $\bar{y}$ et σ_Y^2 de la variable Y sachant que :

$$Y_i = \frac{X_i - b}{a} \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{k} n_iY_i \qquad \sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{k} n_i{Y_i}^2 - \bar{y}^2$$

On déduira la moyenne $\bar{x}$ et la variance σ_X^2 de X.

Exemple : (taille de 50 élèves). On a $a = 3$ et $b = 162.5$

Classes	X_i	n_i	n_{ic}	$\boldsymbol{n_ix_i}$	$\boldsymbol{n_i{x_i}^2}$	$\boldsymbol{Y_i}$	$\boldsymbol{n_iY_i}$	$\boldsymbol{n_i{Y_i}^2}$
[152, 155[	153.5	8	8	1228	188498	-3	-24	72
[155, 158[	156.5	10	18	1565	244922.5	-2	-20	40
[158, 161[	159.5	9	27	1435.5	-	-1	-9	9
[161, 164[	162.5	6	33	975	-	0	0	0
[164, 167[	165.5	6	39	993	-	1	6	6
[167, 170[	168.5	6	45	1011	-	2	12	24
[170, 173[	171.5	5	50	857.5	-	3	15	45
Σ	/	50	/	8065	-	/	-20	196

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{k} n_iY_i = \frac{-20}{50} = -0.4 \qquad \sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{k} n_i{Y_i}^2 - \bar{y}^2 = \frac{196}{50} - (0.4)^2 = 3.76$$

$$\bar{x} = a\bar{y} + b = 3(-0.4) + 162.5 = 161.3$$

$$\sigma_X^2 = a^2\sigma_Y^2 = 9 \times 3.76 = 33.84$$

4/ Coefficient de variation :

C'est le nombre $Cv = \frac{\sigma}{\bar{x}}$ qui permet de relativiser l'écart-type en fonction de la taille des valeurs. Il permet ainsi de comparer la dispersion de séries de mesures exprimées dans des unités différentes car il n'a pas d'unité.

La série avec le plus petit coefficient de variation serait la moins dispersée, c'est-à-dire elle aurait ses valeurs situées plus autour de la moyenne que les autres.

Exemple : (taille de 50 élèves).

$\bar{x}$=161.3 et $\sigma = \sqrt{33.84} = 5.82$ alors $Cv = \frac{\sigma}{\bar{x}} = \frac{5.82}{161.3} = 0.036$

VI/ Paramètres de forme :

1/ Coefficient d'asymétrie de Fisher :

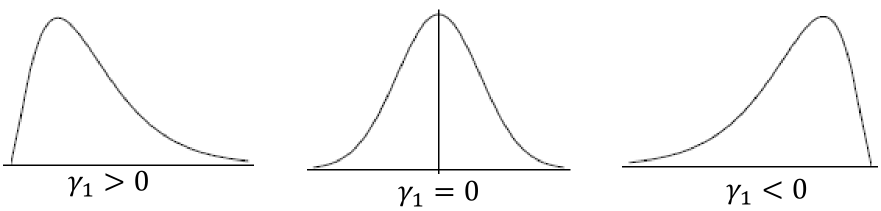
On peut juger visuellement, si une distribution est plus étalée sur la droite ou sur la gauche à partir d'un diagramme en bâtons ou d'un histogramme. Le coefficient d'asymétrie de Fisher est une caractéristique qui permet de mesurer l'asymétrie d'une distribution : $\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$

C'est un nombre sans dimension (indépendant des unités de mesure des x_i).

Si $\gamma_1 = 0$, la distribution est symétrique autour de la moyenne $\bar{x}$.

Si $\gamma_1 > 0$, la distribution est plus étalée sur la droite.

Si $\gamma_1 < 0$, la distribution est plus étalée sur la gauche.



Remarque : en comparant les trois paramètres de position ; mode, médiane et moyenne on peut dire si la série est symétrique on non. On distingue les trois cas suivants :

Si $M_o = M_e = \bar{x}$, la distribution est symétrique.

Si $M_o < M_e < \bar{x}$, la distribution est plus étalée sur la droite.

Si $\bar{x} < M_e < M_o$, la distribution est plus étalée sur la gauche.

Exemple : (taille de 50 élèves).

$M_o = 157 < M_e = 160.33 < \bar{x} = 161.3$ alors la distribution est plus étalée sur la droite (voir graphe).

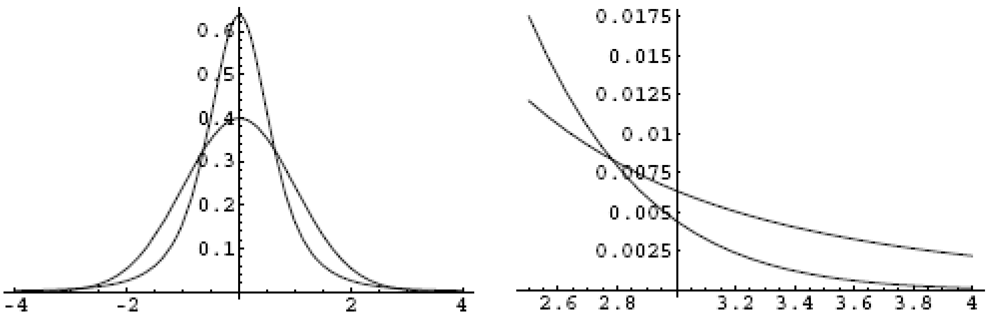
2/ Coefficient d'aplatissement de Fisher :

C'est le nombre (sans dimension) suivant : $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3$

Pour une variable de distribution Normale connue en statistique, $\gamma_2 = 0$. Une telle distribution (en cloche) est souvent considérée comme idéale, d'où :

- Si $\gamma_2 = 0$, la série est normale.
- Si $\gamma_2 > 0$, la série est moins aplatie qu'une série statistique normale de même moyenne et de même variance. La courbe est plus pointue et possède des queues plus longues.
- Si $\gamma_2 < 0$, la série est plus aplatie qu'une série statistique normale de même moyenne et de même variance. La courbe est plus arrondie et possède des queues plus courtes.

Exemple de deux distributions de même moyenne et de même variance. La distribution la plus pointue a une queue plus épaisse.



Exemple : (taille de 50 élèves).