OFPPT

مكتب التكوين المهني وإنعاش الشغل

Office de la Formation Professionnelle et de la Promotion du Travail

Module : Statistiques

Chapitre : Les paramètres de position

Nous avons vu que les représentations graphiques sous forme de tableaux ou sous forme de courbes donnent déjà une première analyse des phénomènes étudiés. Cependant il est plus commode de caractériser une série statistique au moyen d'un nombre représentatif de l'ensemble du phénomène.

Cependant le choix d'une valeur centrale, par exemple la moyenne peut avoir des avantages et des inconvénients : ainsi la moyenne arithmétique permet de caractériser le niveau général d'un étudiant, mais elle masque en même temps son niveau général réel dans chaque discipline.

1- Moyenne arithmétique :

La moyenne arithmétique d'une série statistique χ_1, χ_2, …χ_p est égale à la somme des valeurs observées, divisée par le nombre d'observation. On la note :

$$\bar{\chi} = \frac{\sum_{i=1}^{i=p} \chi_i}{N}, \text{ avec } n = \sum_{i=1}^{i=p} n_i.$$

Remarque: $\bar{\chi}$ peut s'écrire aussi en fonction des fréquences, en effet:

$$\bar{\chi} = \sum_{i=1}^{i=n} \frac{\chi_i}{n} = \sum_{i=1}^{i=n} f_i.$$

Si la série est pondérée, on écrit :

$$\overline{\chi} = \frac{\sum_{i=1}^{i=p} n_i\chi_i}{\sum_{i=1}^{i=p} n_i}, \text{ avec } \frac{\sum_{i=1}^{i=p} n_i\chi_i}{N} = \sum_{i=1}^{=p} \frac{n_i}{N}.\ \chi_i = \sum_{i=1}^{i=p} f_i\chi_i.$$

où $N = \sum_{i=1}^{i=p} n_i$. $\imath_1, n_2, \ldots n_p$ sont les effectifs correspondants à la série observée).

et $f_1, f_2, \ldots F_p$ sont les fréquences correspondants à la série observée.

Exemple :

Les performances en jet javelot de 100 joueurs sont présentées dans le tableau suivant :

Longueur en mètres	71	74	77	80	83
Effectifs	6	17	41	27	9

$$\overline{\chi} = \frac{71 \text{ x } 6 + 74 \text{ x } 17 + 77 \text{ x } 41 + 80\ 27 + 83 \text{ x } 9}{6 + 17 + 41 + 27 + 9} = \frac{7\ 748}{100} = 77{,}48.$$

$\overline{\chi} = 77{,}48.$

Si X est une variable continue de distribution $\{(\lfloor \alpha_i\alpha_{i,+1} \rfloor, n_i) / 1 \leq i \leq p\}$, la moyenne se calcule comme précédemment, en remplaçant χ_i par le centre c_i de la classe

$[\alpha_i, \alpha_{i+1}[$ $\left[c_i = \frac{\alpha_i + \alpha_{i+1}}{2} \right]$.

Exemple :

Soit le tableau donnant les salaires des cadres d'une entreprise et leur fréquence :

χ_i	f_i
[5 000,7 000[	0,21
[7 000,9 000[	0,34
[9 000, 14 000[	0,25
[14 000, 20 000[	0,15
[20 000, 30 000[	0,05
Total	1,00

$\left[c_i = \frac{\alpha_i + \alpha_{i+1}}{2} \right]$.

Par exemple : $c_1 = \dfrac{5\,000 + 7\,000}{2} = 6\,000,$

c_1	f_i
6 000	0,21
8 000	0,34
11 500	0,25
17 000	0,15
25 000	0,05
	1,00

$\overline{\chi} = \sum_{i=1}^{i=5} c_i f_i = 10655.$

2- Moyenne géométrique :

Définition :

La moyenne géométrique notée G ou Mg de n observations, positives est égale à la racine nième de leur produit. Soit χ_1, χ_2, $\ldots\chi_n$ une suite d'observation, alors $G = \sqrt[n]{\chi_1 \text{ x } \ldots \text{ x}\chi_i, \ldots\text{x } \chi_n}$.

Remarque, on peut noter la moyenne géométrique aussi :

$$G = (\chi_1 \text{ x } \ldots \text{ x}\chi_i, \ldots\text{x } \chi_n)^{\frac{1}{n}} = \left[\prod_{i=1}^{i=n} \chi_i\right]^{\frac{1}{n}}$$

Le calcul de G peut s'effectuer grâce à l'utilisation de la fonction logarithme, en effets :

$$\text{Log } G = \log(\chi_1 \text{ x } \ldots \text{ x}\chi_i, \ldots\text{x } \chi_n)^{\frac{1}{n}}$$

$$= \frac{1}{n} \log (\chi_1 \text{ x } \ldots \text{ x}\chi_i, \ldots\text{x } \chi_n) = \frac{\log\chi_1 + \ldots \log \chi_n}{n} = \frac{\sum_{i=1}^{n} \log \chi_i}{n}.$$

Si la série est pondérée, on écrit alors:

$$G = \sqrt[n]{\chi_1^{n_1} \text{ x } \ldots \text{ x}\chi_i^{n_i}, \ldots\text{x } \chi_p^{n_p}}$$

avec $n = \sum_{i=1}^{i=p} n_i = n_1 + \dots n_p$

soit

$$G = (x_1^{n_1} \times \dots \times x_i^{n_i}, \dots \times x_p^{n_p})^{\frac{1}{\sum_{i=1}^{i=p} n_i}} = \left[\prod_{i=1}^{i=p} x_i^{n_i}\right]^{\frac{1}{\sum_{i=1}^{i=p} n_i}} = (x_1^{n_1} \dots x_p^{n_p})^{\frac{1}{n}} = x_1^{\frac{n_1}{n}} \dots x_p^{\frac{n_p}{n}}$$

$$= x_1^{f_1} \dots x_p^{f_p} = \prod_{i=1}^{i=p} x_i^{f_i}$$

et en utilisant le logarithme ,on aura

$$\log G = \ln(x_1^{n_1} \times \dots \times x_i^{n_i}, \dots \times x_p^{n_p})^{\frac{1}{\sum_{i=1}^{i=p} n_i}}$$

$$= \frac{1}{\sum_{i=1}^{i=p} n_i} \ln(x_1^{n_1} \dots x_p^{n_p}) = \frac{n_1 \log x_1 + \dots n_p \log x_p}{\sum_{i=1}^{i=p} n_i}$$

$$\log G = \frac{\sum_{i=1}^{i=p} n_i \log x_i}{\sum_{i=1}^{i=p} n_i} \text{ pour } i = 1 \dots . p.$$

Ainsi la moyenne géométrique des deux nombres réels 2 et 8 est G = $\sqrt{2 \text{ x } 8} = 4$.

Alors que la moyenne arithmétique est $\bar{\chi} = 5$.

Dans le cas où la série est pondérée, il faudra tenir compte des coefficients de pondération. Soit k observations distinctes $\chi_1, \chi_2, \dots \chi_k$, de coefficients respectifs $n_1, n_2, \dots n_k$. Avec n est l'effectif total, soit $n = \sum_{i=1}^{i=k} n_i = n_1 + n_2 + \dots n_k$,

La moyenne géométrique s'écrit d'écrit :

$$G = \sqrt[n]{\chi_1^{n_1} \text{ x } \dots \text{ x} \chi_i^{n_i}, \dots \text{x } \chi_k^{n_k}}$$

Exemple:

Calculer la moyenne géométrique de la série suivante:

χ_i	n_i
2	9
3	3
total	12

En effet : $\sqrt[12]{2^9 \times 3^3} = 2{,}21$.

Alors que la moyenne arithmétique est $\chi = 2.25$ nous remarquons que la moyenne géométrique est inférieure à la moyenne arithmétique.

La moyenne géométrique est adaptée aux phénomènes de croissance, elle est aussi utilisée dans les finances.

3- Moyenne harmonique :

La moyenne harmonique H d'une série statistique est égale à l'inverse de la moyenne arithmétique des inverses des valeurs observées.

Si n est le nombre d'observations $\chi_1, \chi_2, \ldots \chi_n$, la moyenne harmonique est :

$$H = \frac{N}{\sum_{i=1}^{i=p} \frac{1}{x_i}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{i=p} \frac{1}{x_i}}.$$

Si la série est pondérée, on écrit:

$$H = \frac{\sum_{i=1}^{i=p} n_i}{\sum_{i=1}^{i=p} \frac{n_i}{x_i}} = \frac{1}{\frac{1}{N}\sum_{i=1}^{i=p} \frac{n_i}{x_i}} = \frac{1}{\sum_{i=1}^{i=p} \frac{n_i}{N} \cdot \frac{1}{x_i}} = \frac{1}{\sum_{i=1}^{i=p} f_i \cdot \frac{1}{x_i}}.$$

Soit $H = \frac{1}{\sum_{i=1}^{i=p} \frac{f_i}{x_i}}$.

Exemple :

Un escargot monte un puit de longueur trente mètres en une heure, et en descendant sa vitesse double. Quelle est la vitesse moyenne de cet escargot ?

En effet une approche intuitive conduit à dire que sa vitesse moyenne est

$v_{moy} = \frac{30+60}{2} = 45$ mètres à l'heure. Ce qui est manifestement faux. En effet une vitesse

est une grandeur qui dépend de deux paramètres, la distance et le temps et dans notre exemple, la distance du puit étant constante, et ce sont les temps qui varient. Soit *d* cette distance, v_1 et v_2 les deux vitesses, t_1 et t_2 les temps de parcours respectifs à la montée et la décente.

En montant $v_1 = \frac{D}{t_1}$ et en descendant $v_2 = \frac{d}{t_2}$, donc,

$$v_{moy} = \frac{d + d}{t_1 + t_2} = \frac{2d}{\frac{d}{v_1} + \frac{d}{v_2}} = \frac{2}{\frac{1}{v_1} + \frac{1}{v_2}} .$$

Nous remarquons donc que la vitesse moyenne est ici la moyenne harmonique des vitesses.

Soit $v_{moy} = \frac{2}{\frac{1}{30} + \frac{1}{60}} = 40$ mètres/h.

L'usage de la moyenne harmonique est peu fréquent, cependant elle est utilisée dans les calculs des indices.

4- Moyenne quadratique :

La moyenne quadratique d'une série statistique est la racine carrée de la moyenne arithmétique des carrées des valeurs observées. On la note

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{i=p} x_i^2} .$$

Si la série est pondérée, c'est-à-dire affectée par des coefficients de pondération, on écrit alors :

$$Q = \sqrt{\frac{\sum_{i=1}^{i=p} n_i x_i^2}{\sum_{i=1}^{p} n_i}} = \sqrt{\sum_{i=1}^{p} \frac{n_i}{n} \cdot x_i^2} = \sqrt{\sum_{i=1}^{p} f_i x_i^2} ,$$

On note aussi

$$Q=\left[\frac{\sum_{i=1}^{p} n_i x_i^2}{\sum_{i=1}^{p} n_i}\right]^{\frac{1}{2}}=\left[\sum_{i=1}^{p} f_i x_i^2\right]^{\frac{1}{2}}.$$

Exemple :

Soit la distribution de 10 groupes industriels selon le nombre de filiales délocalisées à l'extérieur du pays d'origine.

Nombre de groupes	Nombres de filiales
1	5
2	15
3	10
4	10

Calculons les valeurs des moyennes arithmétique, harmonique, quadratique et géométrique et vérifions leurs positions respectives.

Nombre de groupes x_i	Nombre de filiales n_i	$n_i\ x_i$	n_i / x_i	$\log x_i$	$n_i \log x_i$	x_i^2	$n_i\ x_i^2$
1	5	5	5	0	0	1	5
2	15	30	7,5	0,30103	4,51545	4	60
3	10	30	3,33	0,47712	4,7712	9	90
4	10	40	2,5	0,60206	6,0206	16	160
Total	40	105	18,33		15,30725		315

La moyenne arithmétique :

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{i=4} n_i x_i=\frac{105}{40}=2{,}625.$$

La moyenne harmonique:

$$H=\frac{n}{\sum_{i=1}^{i=4}\frac{n_i}{x_i}}=\frac{40}{18{,}33}=2{,}182.$$

La moyenne géométrique:

$$Q = \sqrt{\frac{1}{n}\sum_i n_i x_i^2} = \sqrt{\frac{315}{40}} = \sqrt{7,875} = 2,806.$$

La moyenne quadratique:

$$\log G = \frac{1}{n}\sum_{i=1}^{i=4} n_i \log x_i = \frac{15,30725}{40} = 0,38268$$

d'où $G = 2,414.$

Les valeurs des moyennes s'échelonnent ainsi:

$2,182 < 2,414 < 2,685 < 2,806.$

On obtient: $H < G < \bar{x} < Q.$

Remarque :

Si une série statistique pour laquelle les quatre moyennes définies ci-dessus existent, on a alors H < G < $\bar{\chi}$ < Q. Cette remarque est très utile dans les exercices, elle sert de moyen de vérification de l'exactitude des calculs effectués.

5- Le mode :

Définition :

Si X est une variable discrète, on appelle mode toute valeur x_i dont l'effectif (ou la fréquence) est maximum. Si X est une variable continue, on appelle classe modale

X $[\alpha_i, \alpha_{i+1}[$ pour laquelle $\frac{n_i}{\alpha_{i+1} - \alpha_i}$ (ou $\frac{f_i}{\alpha_{i+1} - \alpha_i}$) est maximum (sur l'histogramme, est maximum).

Le mode permet de connaître la valeur la plus probable du caractère.

Cette valeur centrale est simple à percevoir. Mais elle ne tient pas compte de l'ensemble des valeurs du caractère de la série étudiée. Elle ne s'intéresse qu'aux effectifs (ou fréquences) des différentes modalités.

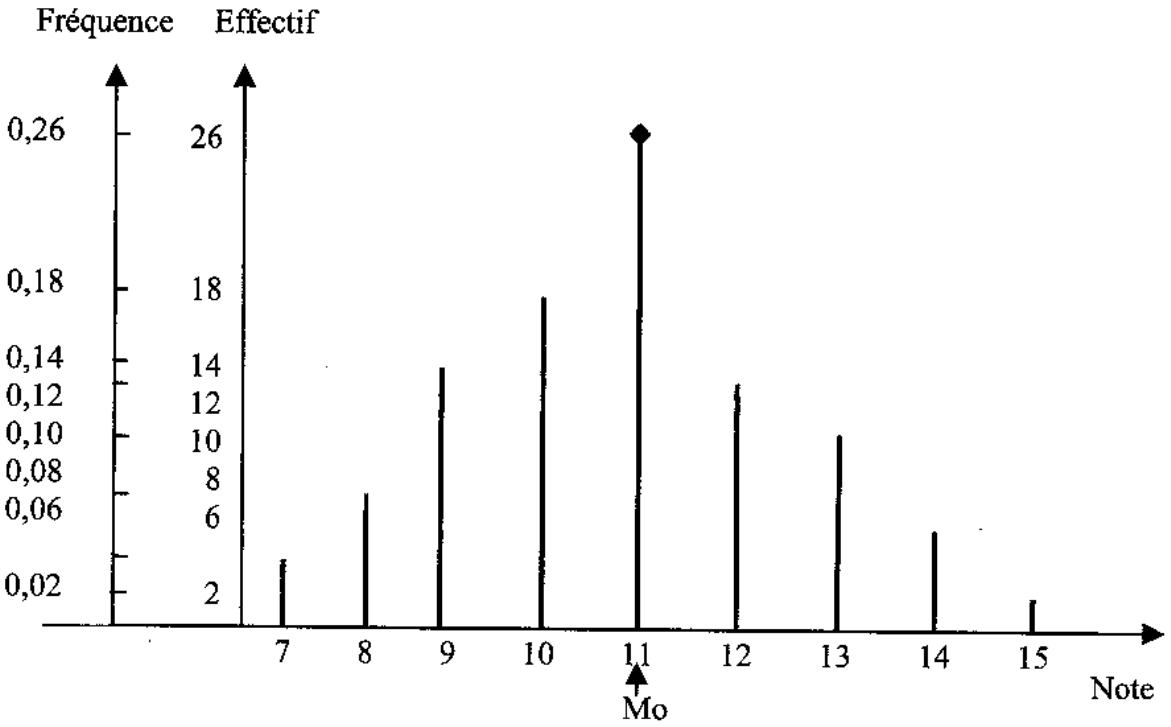
Séries à caractère qualitatif ou à caractère quantitatif discret (discontinu)

Pour de telles séries, la détermination du mode est immédiate.

Exemple :

Les notes obtenues par 100 étudiants dans un examen sont indiquées dans le tableau ci-dessous :

Notes	Effectifs	Fréquences
7	4	0,04
8	7	0,07
9	14	0,14
10	18	0,18
11	26	0,26
12	13	0,13
13	10	0,10
14	6	0,06
15	2	0,02
Total	100	1,00



Série à caractère continu

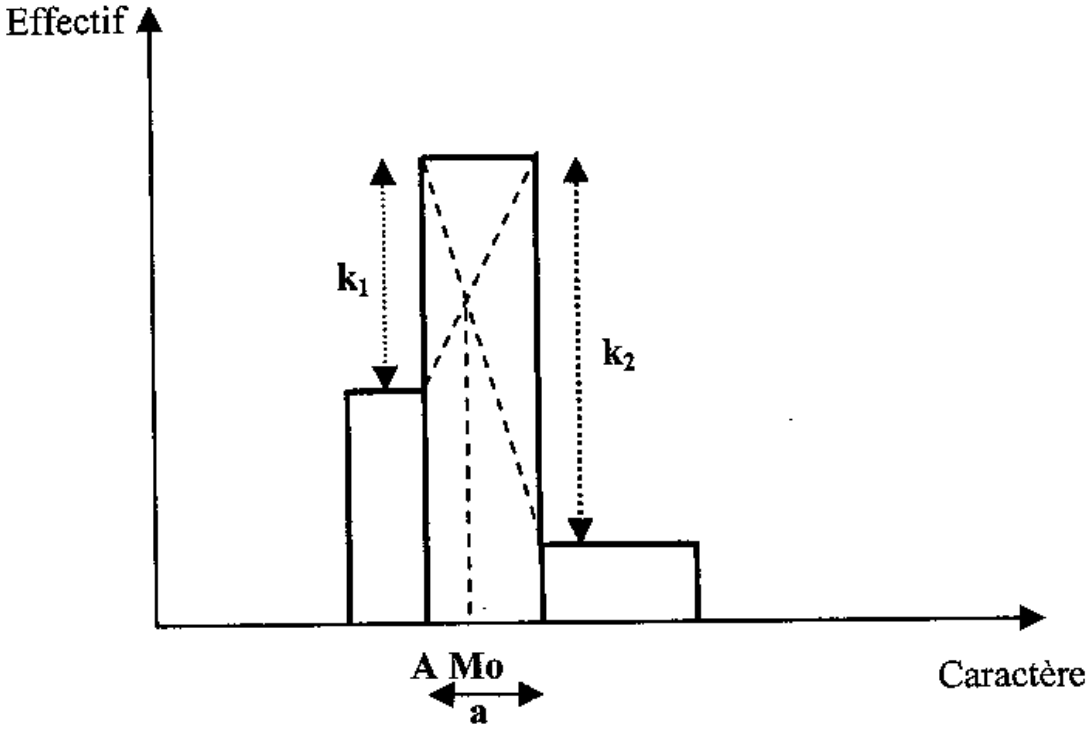
La détermination du mode est plus délicate pour les séries à caractère continu.

On perçoit immédiatement la classe correspondant à la fréquence la plus grande ou à l'effectif le plus important. Cette classe porte le nom de classe modale ; elle contient le mode.

Pour déterminer le mode, plusieurs procédés sont possibles.

-Si on ne s'intéresse pas à une valeur plus précise, il est possible de considérer que le mode de la série correspond au milieu de la classe modale.

-Si au contraire, on désire plus de précision, le procédé le plus simple est un procédé graphique. Il consiste à estimer que le mode est déporté à l'intérieur de la classe modale en fonction des effectifs rectifié des classes encadrants la classe modale. Le mode est alors obtenu de la façon indiquée sur la figure ci-dessous.



Série continue
Détermination graphique du mode

-de manière encore plus précise, on peut calculer le mode par la formule suivante, qui correspond à la traduction mathématique de la méthode graphique :

$$Mo = A + \frac{k_1}{k_1 + k_2} \times \alpha$$

A est la valeur du caractère correspondant à la valeur inférieur de classe modale ; **a** est l'amplitude de la classe modale ; k_1 et k_2 sont les différences entre les effectifs (ou les fréquences) rectifiés de la classe modale et des deux classes qui l'encadrent.

Exemple :

Une étude portant sur la durée de vie d'une espèce biologique aquatique de la même race est consignée dans le tableau ci-dessous :

Durée de vie en heures	Nombre d'espèces
[0,2 000[	8
[2 000, 4 000[	26
[4 000, 5 000[	20
[5 000,6 000[	22
[6 000, 8 000 [	18
[8 000,10 000[	6

Déterminons la classe modale, et plus précisément le mode ;

Les classes ne sont pas d'amplitude égales ; la classe modale n'est donc pas nécessairement celle où l'effectif est le plus élevé. En effet, la classe [2 000, 4 000[dont l'effectif est 26 a une amplitude double de la classe [5 000,6 000[dont l'effectif est 22 ;

La classe dont la densité est maximum, est la classe modale. La densité est calculée à l'aide des effectifs et de l'amplitude. On a :

Durée de vie en heures	Nombre d'espèces n_i	Amplitude a_i	Densité d_i
[0,2 000[	8	2 000	0,004
[2 000, 4 000[	26	2 000	0,013
[4 000, 5 000[	20	1 000	0,020
[5 000,6 000[	22	1 000	0,022
[6 000, 8 000 [	18	2 000	0,009
[8 000,10 000[	6	2 000	0,003

$d_i = \frac{n_i}{a_i}$, on obtient alors le tableau suivant :

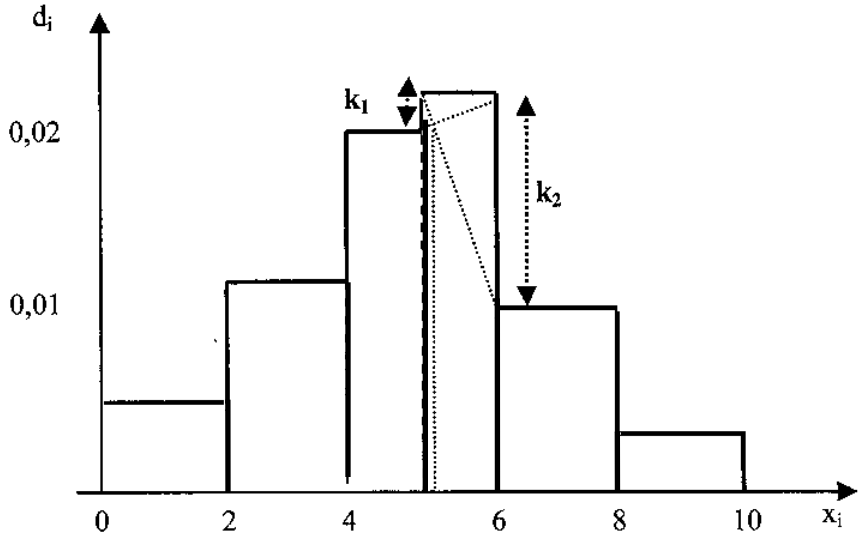
La classe modale est donc [5,6[.

En appliquant la méthode proposée ci-dessus, on aura :

$$Mo = A + \frac{k_1}{k_1 + k_2} \times a,$$

$$Mo = 5\,000 + \frac{0{,}002}{0{,}002 + 0{,}013} \times 1000 = 5\,133{,}33$$

Résolution graphique



Enfin le mode peut ne pas être unique. On parle alors de distribution multimodale ou (plurimodale).

6- Médiane :

Définition :

Si X est une variable discrète prenant n valeurs : $v_1 \leq v_2 \leq \ldots v_n$ (les valeurs prises par X ne sont pas groupées), on appelle médiane un nombre réel Me tel qu'il y ait autant de valeurs v_j inférieurs ou égales à Me que de valeurs v_j supérieures ou égales à Me.
Si n = 2k + 1, c'est-à-dire n est impair $m_e = v_{k+1}$.

Si n = 2k, c'est-à-dire n est pair par convention $Me = \frac{vk + v_{k+1},}{2}$.

Si X est une variable continue, on appelle *médiane* le nombre réel Me abscisse du point d'ordonnée $\frac{1}{2}$ de la courbe cumulative des fréquences, c'est-à-dire le nombre réel Solution de l'équation $Fx(t) = \frac{1}{2}$.

Me appartient à la première classe $[\alpha_i\alpha_{i+1}[$ (dont la fréquence cumulée p_i est supérieure ou égale à 0,5 (on a donc : $p_{i-1} < 0,5$ et $p_i \geq 0,5$).

La valeur de Me s'obtient en résolvant : $\frac{Me - a_i}{a_{i+1} - a_i} = \frac{0,5 - p_{i-1}}{p_i - p_{i-1}}$, donc

$Me = a_{i+}\frac{0,5 - p_{i-1}}{P_i - p_{i-1}} (a_{i+1} - a_i)$.

La courbe cumulative des fréquence fournit graphiquement une valeur approchée de Me.

Cas des séries à caractère discret

Exemple 1 :

Les performances en jet de javelot de 100 joueurs présentées dans le tableau suivant :

Longueur en m	71	74	77	80	83
Effectifs	6	17	41	27	9

Cette distribution peut se présenter de la manière suivante :

$$\underbrace{71...71}_{6}\ \underbrace{74...74}_{17}\ \underbrace{77...77}_{41}\ \underbrace{80...80}_{27}\ \underbrace{83...83}_{9}$$

$$n = 100 = 2 \times 50 \text{ donc } Me = \frac{v_{50} + v_{51}}{2} = \frac{77 + 77}{2} = 77.$$

Exemple 2 :

Considérons une classe de 60 élèves qui ont eux les notes suivantes dans un examen de statistique :

Not χ_i	4	5	5,5	7	8,5	9	10	12	13
Effectif n_i	5	6	3	8	10	12	10	4	2
Effectif cumulé	5	11	14	22	32	44	54	58	60

La médiane est 8,5 car il y a autant de valeur v_j inférieures ou égales à Me que de valeurs v_j supérieures ou égales à Me.

Cas des séries à caractère continu

Dans ce cas la valeur de la médiane du caractère appartient à une classe. La médiane est déterminée, soit par le calcul, soit par le graphique.

Exemple :

Considérons le tableau suivant donnant la superficie de 100 parcelles de terrains agricoles en hectares.

Superficie χ(en ha)	$20 \le x < 40$	$40 \le x < 60$	$60 \le x < 80$	$80 \le x < 100$	$100 \le x < 120$
Effectif	12	31	32	15	10
Effectif cumulé	12	43	75	90	100

a) résolution par calcul :

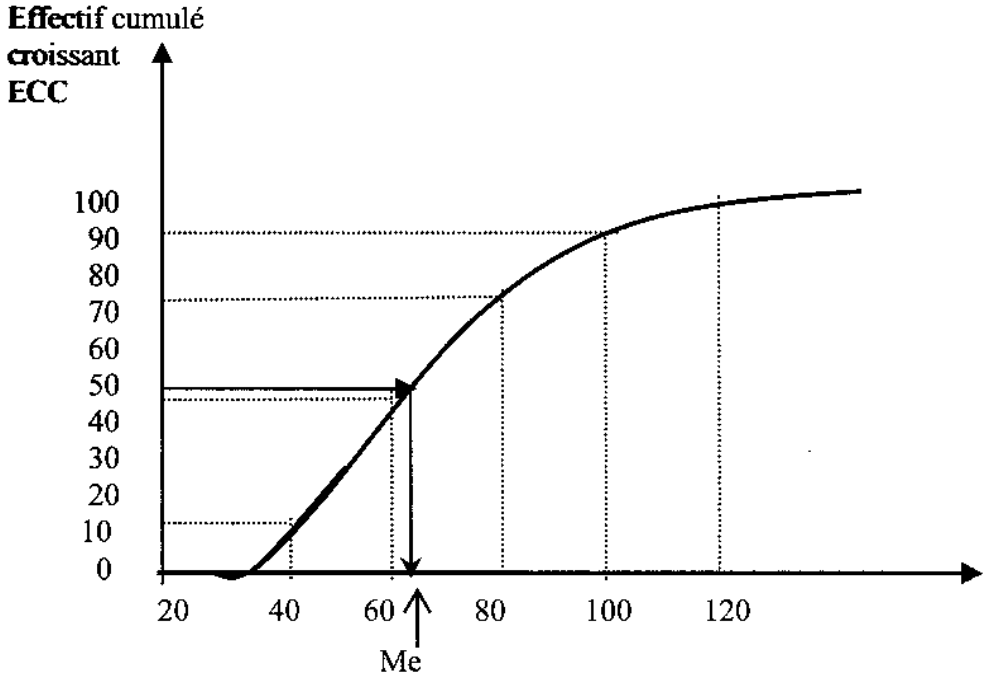
La médiane Me correspond à l'effectif cumulé 50, donc elle appartient à la classe $60 \le x < 80$.

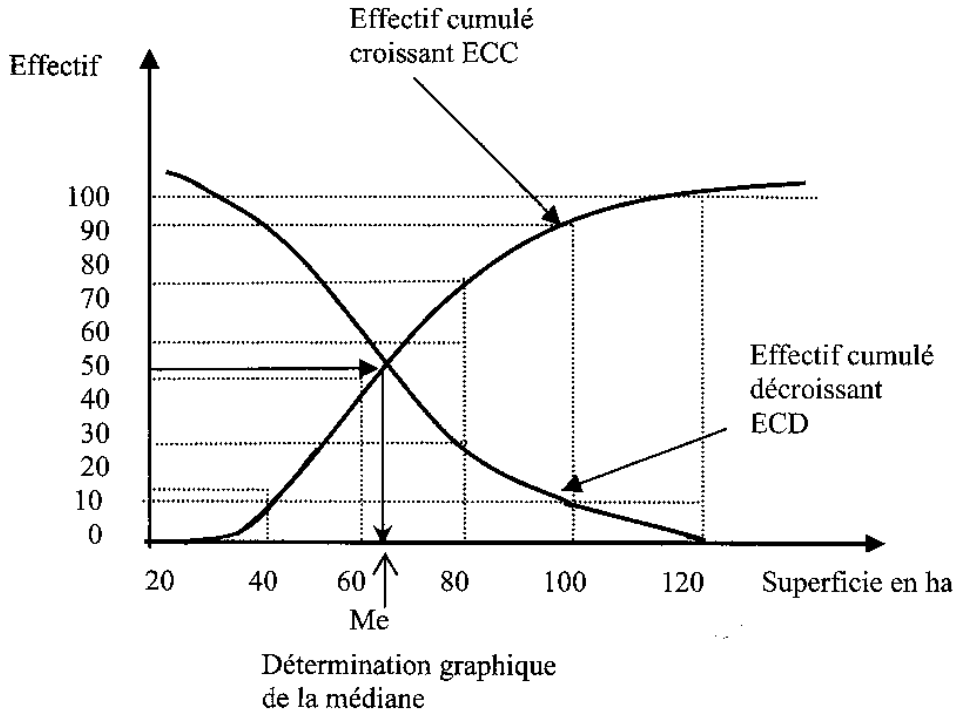
Par la méthode d'interpolation linéaire on a :

$$\frac{Me - 60}{50 - 43} = \frac{80 - 60}{75 - 43},$$

Cela correspond à $Me = 60 + (50 - 43) \times \frac{80 - 60}{75 - 43} = 64{,}375$.

b) résolution graphique :





La médiane est le point d'intersection des deux courbes cumulatives.