

Les caractéristiques de tendance centrales nous permettent d'avoir un ordre de grandeurs de la série mais ne nous renseignent pas sur la structure interne de la série ainsi par exemple les trois séries suivantes :

- 58 ; 59 ; 60 ; 61 ; 62 ; dont $\bar{x} = 60$
- 50 ; 55 ; 60 ; 65 ; 70 ; dont $\bar{x} = 60$
- 1 ; 1 ; 1 ; 1 ; 296 ; dont $\bar{x} = 60$

Elles ont le même ordre de grandeurs : $N = 5$ et la même moyenne $\bar{x} = 60$ mais elles n'ont pas la même structure interne ; dans la troisième série les valeurs du caractère sont beaucoup plus dispersées que dans la première et la deuxième d'où l'intérêt d'un indicateur de dispersion.

I) L'INTERVALLE DE VARIATION OU L'ETENDUE :

L'étendue (e) est la différence entre la plus grande et la plus petite des valeurs observées. Ce paramètre est également appelé « intervalle de variation ».

- Cette caractéristique est la plus simple mais aussi la moins significative.
- Sa signification est claire et son calcul extrêmement rapide.
- Ces avantages la font fréquemment utiliser dans le contrôle de fabrication industrielle plutôt que d'effectuer des calculs complexes en atelier.

Exemple : dans une série statistique on 'a relevé les informations suivantes :

25	40	45	60	70	75	80
----	----	----	----	----	----	----

L'intervalle de variation :

$$e = 80 - 25 = 55$$

Signification :

Ce calcul est simple mais la simplicité de ce calcul ne doit pas nous faire oublier que « l'étendue » est très sensible aux fluctuations des valeurs « extrêmes » qui sont souvent peu représentatives.

Cette valeur caractéristique, qui correspond à un concept fort utilisé dans la pratique (écart entre le premier et le dernier coureur, écart entre la meilleur et la plus faible note, etc.) est insuffisante pour une étude sérieuse de la dispersion.

II) L'ECART INTERQUATILE :

L'intervalle interquartile permet de supprimer les valeurs erronées éventuelles qui se trouvent en extrémité d'une distribution.

A) DEFINITION DU QUARTILE :

La définition du quartile est analogue, dans son principe, à celle de la médiane. Il y a trois quartiles : Q_1 ; Q_2 et Q_3 .

Ce sont les valeurs de la variable statistique telles que, les observations étant rangées en ordre croissant, un quart des observations soient inférieures à Q_1 , un quart comprises entre Q_1 et Q_2 , un quart comprises entre Q_2 et Q_3 et un quart supérieures à Q_3 .

- On appelle premier quartile Q_1 : la valeur du caractère telle que 25% des effectifs ont une valeur inférieure à Q_1 et 75% des effectifs lui sont supérieures

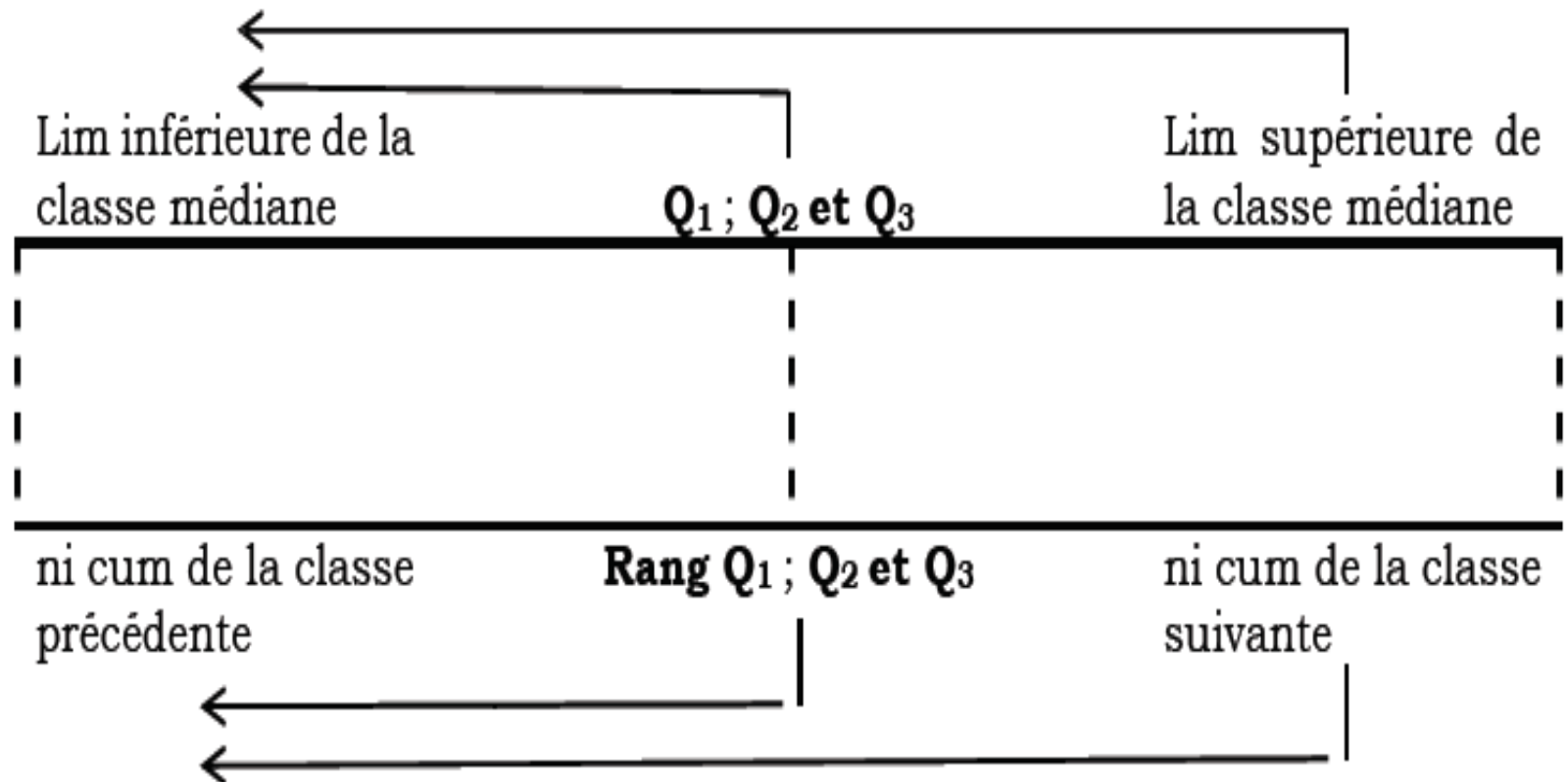
CARACTERISTIQUES DE DISPERSION Page 5

- On appelle deuxième quartile Q_2 : la valeur du caractère telle que 50% des effectifs ont une valeur inférieure à Q_2 et 50% des effectifs lui sont supérieures : c'est la médiane (**Me**).
- On appelle troisième quartile Q_3 : la valeur du caractère telle que 75% des effectifs ont une valeur inférieure à Q_3 et 25% des effectifs lui sont supérieures.

Pour calculer les quartiles, il faut suivre les étapes suivantes :

- 1) Calculer les effectifs cumulés croissants ou décroissants,
- 2) Calculer le rang du:
 - Premier quartile : rang $Q_1 = \sum n_i / 4$
 - Deuxième quartile : rang $Q_2 =$ rang de la Médiane $= \sum n_i / 2$
 - Troisième quartile : rang $Q_3 = 3 \sum n_i / 4$,
- 3) Repérer les rangs de Q_1 ; Q_2 et Q_3 dans la colonne des n_i cumulés croissants ou décroissants,
- 4) Déterminer les classes de Q_1 ; Q_2 et Q_3 .
- 5) Interpoler linéairement pour situer exactement la valeur de Q_1 ; Q_2 et Q_3 .

CARACTERISTIQUES DE DISPERSION Page 6



B) APPLICATIONS :

CARACTERISTIQUES DE DISPERSION Page 7

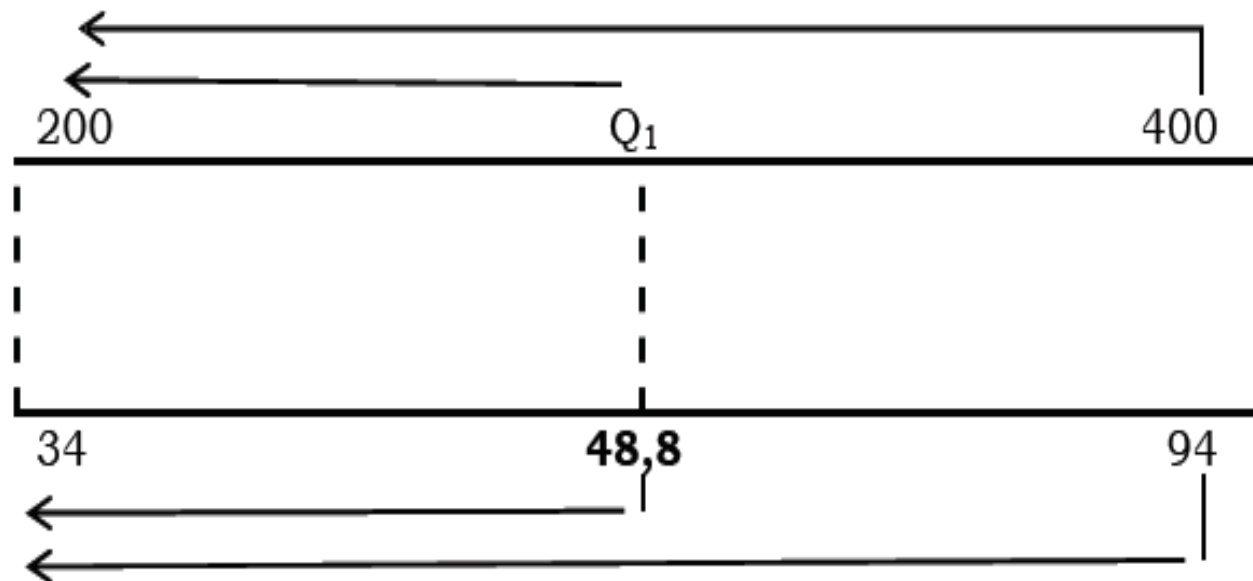
a) Calculons le premier quartile des 195 employés de l'entreprise industrielle

Primes mensuelles		Employés n_i	n_i cumulés croissants	
[0 - 200]	Q₁	34	34	97,5
[200 - 400]		60	94	
[400 - 600]		23	117	
[600 - 800]		30	147	
[800 - 1000]		24	171	
[1000 - 1200]		13	184	
[1200 - 1400]		11	195	
Totaux		195		

1) Calculer les effectifs cumulés croissants : voir tableau ci-dessus,

CARACTERISTIQUES DE DISPERSION Page 8

- 2) Calculer le rang de Q_1 : $\text{rang } Q_1 = \sum n_i / 4 = 195 / 4 = 48,8$
- 3) Repérer le rang de Q_1 dans la colonne des n_i cumulés croissants ou décroissants,
- 4) Déterminer la classe de Q_1 : le rang de Q_1 se situe entre la valeur 34 et 94 car : $34 < 48,8 < 94$ donc la classe de Q_1 est $[200 - 400]$,
- 5) Il faut interpoler linéairement pour situer exactement la valeur de Q_1



$$Q1 = \frac{Q1 - 200}{400 - 200} = \frac{48,8 - 34}{94 - 34}$$

$$Q1 = \frac{Q1 - 200}{200} = \frac{14,8}{60}$$

$$Q1 = 249,16dh$$

Signification :

C'est la prime telle que 75% des employés touchent une prime supérieure à 249,16dh et 25% des employés touchent une prime inférieure à 249,16dh

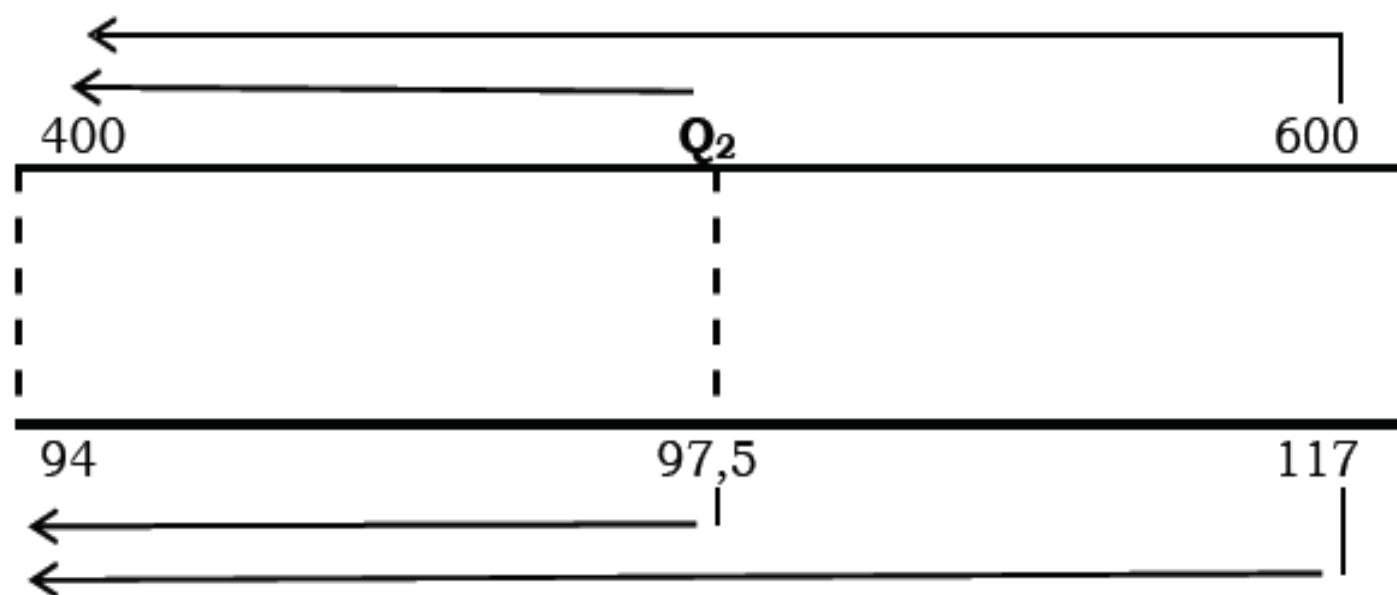
b) Calculons le deuxième quartile des 195 employés de l'entreprise industrielle.

Primes mensuelles	Q ₂	Employés n _i	n _i cumulés croissants	97,5
[0 - 200]		34	34	
[200 - 400]		60	94	
[400 - 600]		23	117	
[600 - 800]		30	147	
[800 - 1000]		24	171	
[1000 - 1200]		13	184	
[1200 - 1400]		11	195	
Totaux		195		

CARACTERISTIQUES DE DISPERSION

Page 11

- 1) Calculer le rang de Q_2 correspond à celui de la médiane :
$$\text{Rang } Q_2 = \sum n_i / 2 = 195 / 2 = 97,5$$
- 2) Déterminer la classe de Q_2 ou médiane : $[400 - 600]$,
- 3) Il faut interpoler linéairement pour situer exactement la valeur de la Q_2 .



$$Q_2 = \frac{Q_2 - 400}{600 - 400} = \frac{97,5 - 94}{117 - 94}$$

$$Q_2 = \frac{Q_2 - 400}{200} = \frac{3,5}{23}$$

$$Q_2 = Me = 430,40dh$$

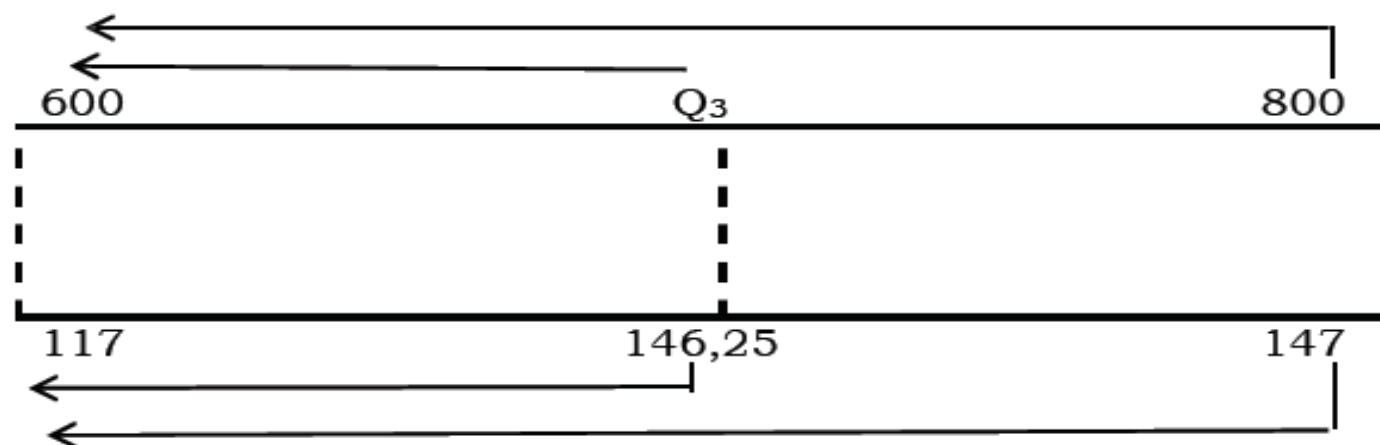
Signification :

C'est la prime telle que 50% des employés touchent une prime supérieure à 430,40dh et 50% des employés touchent une prime inférieure à 430,40dh.

[0 - 200]		34	34
[200 - 400]		60	94
[400 - 600]	Q₃	23	117
[600 - 800]		30	147
[800 - 1000]		24	171
[1000 - 1200]		13	184
[1200 - 1400]		11	195
Totaux		195	

146,25

- 1) Calculer les effectifs cumulés croissants : voir tableau ci-dessus,
- 2) Calculer le rang de $Q_3 = 3 * \sum n_i / 4 = (3 * 195) / 4 = 146,25$
- 3) Repérer le rang de Q_3 dans la colonne des n_i cumulés croissants ou décroissants,
- 4) Déterminer la classe de Q_3 : le rang de Q_3 se situe entre 94 et 117 donc la classe de Q_3 est [400 - 600],
- 5) Il faut interpoler linéairement pour situer exactement la valeur de Q_3 .



$$Q3 = \frac{Q3 - 600}{800 - 600} = \frac{146 - 117}{147 - 117}$$

$$Q3 = \frac{Q3 - 600}{200} = \frac{29,3}{30}$$

$$Q3 = 795$$

Signification :

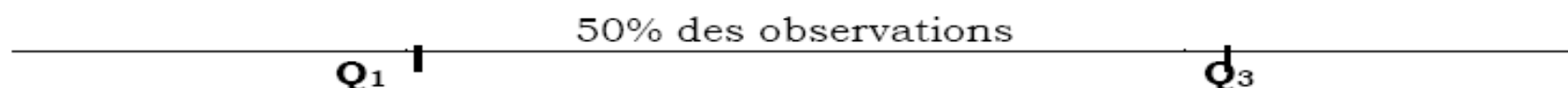
C'est la prime telle que 25% des employés touchent une prime supérieure à 795dh et 75% des employés touchent une prime inférieure à 795dh.

C) DEFINITION DE L'ECART INTERQUARTILE :

CARACTERISTIQUES DE DISPERSION

Page 15

C'est la différence entre le troisième quartile et le premier quartile = $Q_3 - Q_1$
Entre Q_3 et Q_1 . C'est donc l'intervalle qui contient 50 % des observations, en laissant 25 % à gauche et 25 % à droite.



D'après l'exemple que nous avons étudié :

$Q_3 - Q_1 = 795 - 249,16 = 545,84$ dh, cela signifie que 50% des employés ont un écart de prime au plus égal à ce montant.

Les avantages de l'intervalle interquartile résident dans la rapidité de son calcul et la simplicité de sa signification.

Par contre, il ne tient compte que de l'ordre des observations et non de leurs valeurs et des écarts qui existent entre elles.

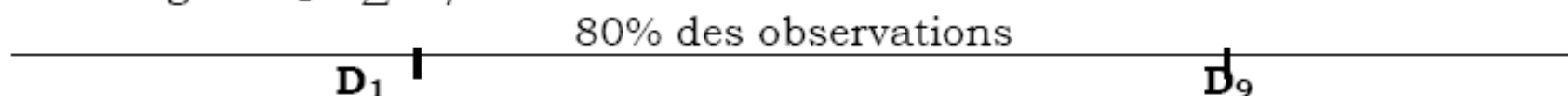
Remarques :

- On utilise quelque fois ce qu'on appelle la déviation quartile ou encore, L'écart semi-interquartile = $\frac{Q_3 - Q_1}{2}$
- Pour comparer des séries de natures différentes, ou exprimées en unités différentes, on utilise l'intervalle interquartile relatif = $\frac{Q_3 - Q_1}{Me}$
- L'inconvénient de l'écart interquartile est de ne tenir compte que de 50% des observations. Certains ont cherché à élargir le nombre d'observations en calculant des écarts inter déciles et des écarts inter percentiles.

D) LES DECILES :

- On appelle premier décile que l'on note D_1 , la valeur du caractère telle que 10% des observations ont une valeur qui est inférieure à D_1 et 90% des observations ont une valeur qui lui est supérieure.

$$\text{Rang de } D_1 = \sum n_i / 10$$



- On appelle le 9ème décile, D_9 : la valeur du caractère tel que 90% des observations lui sont inférieures, et 10% des observations lui sont supérieures.

$$\text{Rang de } D_9 = 9 * \sum n_i / 10$$

- L'intervalle inter décile $D_9 - D_1$: il contient 80% des observations
- Rang de $D_4 = 4 * \sum n_i / 10$
- Rang de $D_5 = 5 * \sum n_i / 10 = \text{rang de la médiane}$
- Rang de $D_8 = 8 * \sum n_i / 10$
-
- Pour les interpolations linéaires on garde toujours le même principe de calcul.

VI) LA VARIANCE ET L'ECART TYPE :

Dans le calcul de l'écart absolu moyen intervenait les valeurs absolues des écarts à la moyenne.

L'écart-type sera défini à partir des valeurs de ces écarts élevés au carré. On déterminera de cette façon une sorte de distance moyenne des observations à la moyenne arithmétique qui constitue une mesure de la dispersion.

A) DEFINITION :

On appelle une variance la moyenne arithmétique des carrés des écarts entre les valeurs du caractère par rapport à leur moyenne arithmétique.

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i}$$

On appelle écart-type est la racine carré positive de la variance

$$\sigma = \sqrt{\sigma^2}$$

B) APPLICATION :

Dans la série suivante on a :

Primes mensuelles	EFFECTIFS n_i	x_i	$n_i \cdot x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
[0 - 200[34	100	3400	-433,85	188225,82	6399677,88
[200 - 400[60	300	18000	-233,85	54685,82	3281149,20
[400 - 600[23	500	11500	-33,85	1145,82	26353,86
[600 - 800[30	700	21000	166,15	27605,82	828174,60
[800 - 1000[24	900	21600	366,15	134065,82	3217579,68
[1000 - 1200[13	1100	14300	566,15	320525,82	4166835,66
[1200 - 1400[11	1300	14300	766,15	586985,82	6456844,02
Totaux	195		104100	1163,05	1313240,74	24376614,90

$$\bar{x} = \frac{\sum(n_i \cdot x_i)}{\sum n_i} = \frac{104100}{195} = 533,85$$

$$\sigma^2 = \frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i} = \frac{24376614,90}{195} = 125008$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{125008} = 353,56$$

Signification :

353,56dh c'est la moyenne des écarts par rapport à la moyenne.

Si la valeur de l'écart type est faible cela veut dire que presque toutes les valeurs de x_i sont proches de la moyenne, on a donc une série peu dispersée. Par contre si l'écart type est élevé cela veut dire qu'il y'a des x_i faibles et des x_i élevées par rapport à la moyenne , on a donc affaire à une série très dispersée.

C) COEFFICIENT DE VARIATION

Supposons que l'on ait calculé la moyenne arithmétique \bar{x} et l'écart-type σ d'une population.

Pour comparer deux séries différentes, on utilise le coefficient de variation

$$v \text{ en } \% = \frac{\sigma}{\bar{x}} * 100$$

Exemple :

Soient les séries suivantes :

Modèle d'ampoules I	Modèle d'ampoules II
$\bar{x} = 1400$ heures	$\bar{x} = 1800$ heures
$\sigma = 100$ heures	$\sigma = 450$ heures
$v = (100/1400) * 100 = 7,14\%$	$v = (450/1800) * 100 = 25\%$

La première série est moins dispersée que la deuxième.

Bien que la notion de concentration soit différente de celle de dispersion, les deux phénomènes varient dans le même sens.

A une grande dispersion des valeurs x_i , correspond obligatoirement une forte concentration.

De plus, les paramètres de concentration qui sont l'écart entre la médiale et la médiane, la courbe de Lorenz et l'indice de Gini, s'interprètent comme les paramètres de dispersion : le degré de concentration est positivement lié à leur valeur.

La notion de concentration n'est applicable que lorsque les valeurs du caractère a une signification

Par exemple :

- Ensemble d'entreprises classées selon le chiffre d'affaires,
- Ensemble de personnes classées selon le revenu,
- Ensemble d'agriculteurs classés selon la surface d'exploitation.

Voici des contres exemples :

- Ensemble d'individus classés selon l'âge ou selon la taille qui n'a pas de sens.

I) LA DETERMINATION ALGEBRIQUE DE LA CONCENTRATION

A) PRINCIPE :

La différence entre la médiale et la médiane, $M_l - M_e$, donne une idée de la concentration d'une distribution.

Pour mesurer cette concentration, on mesure d'abord, l'écart qui existe entre la médiale et la médiane : $\Delta M = M_l - M_e$.

Si $\Delta M = 0$: la concentration est nulle car la répartition est égalitaire ; par contre si $\Delta M \neq 0$: il y'a une concentration.

Lorsque cette différence est grande par rapport à l'étendue, on dira que la concentration est forte, et lorsqu'elle est petite, on dira que la concentration est faible.

Cette concentration est d'autant plus forte que ΔM est grande par rapport à l'intervalle de variation.

$$\text{Concentration : } \frac{M_l - M_e}{\text{L'intervalle de variation}} = \frac{\Delta M}{e}$$

B) APPLICATION :

Soient les salaires horaires des ouvriers d'une entreprise réparties dans le tableau suivant :

Salaires	Ouvriers ni	xi	ni .xi	ni cum croi	ni .xi Cum croi
[20 - 40[8	30	240	8	240
[40 - 50[8	45	360	16	600
[50 - 60[12	55	660	28	1260
[60 - 80[8	70	560	36	1820
[80 - 100[4	90	360	40	2180
	40		2180,00		

La détermination de la concentration nécessite la connaissance de la médiane et la médiale :

a) Calcul de La médiale (Me) :

$$Me = \frac{Me - 50}{60 - 50} = \frac{20 - 16}{28 - 16}$$

$$Me = \frac{Me - 50}{10} = \frac{4}{12}$$

$$Me = 53,33dh$$

b) Calcul de la médiale (Ml) :

$$ML = \frac{ML - 50}{60 - 50} = \frac{1090 - 600}{1260 - 600}$$

$$ML = \frac{ML - 50}{10} = \frac{490}{660}$$

$$ML = 57,42dh$$

c) Mesure de la concentration

$$\Delta M = Ml - Me$$

$$\Delta M = 57,42 - 53,33$$

$$\Delta M = 4,09$$

$$\text{L'intervalle de variation} = e = 100 - 20 = 80$$

$$\text{Concentration} = \frac{Ml - Me}{\text{L'intervalle de variation}} = \frac{\Delta M}{e}$$

$$\text{Concentration} = \frac{4,09}{100 - 20} = \frac{4,09}{80} 0,0511 = 5,11\% : \text{Faible concentration}$$

II) LA DETERMINATION GRAPHIQUE DE LA CONCENTRATION : LA COURBE DE LORENTZ GINI :

A- LA CURBE DE GINI LORENTZ :

GINI propose de mesurer la concentration en mettant en abscisses les fréquences cumulées en%, et en ordonnées x_i cumulés en %

EXEMPLE :

Utilisons la distribution statistique suivante portant sur un ensemble de parcelles de terrains classées d'après leur superficie

Surface (hectares)	Nombre de parcelles
[0 – 10[16
[10 – 20[30
[20 – 40[18
[40 – 70[10
[70 – 100[06
	80

Nous allons étudier le phénomène de concentration et le mesurer.

Surface	Centre de classe x_i	Effectif n_i	Effectifs cumulés croissants	% des Fréquence cumulée croissante	$x_i n_i$	$x_i n_i$ cumulés croissants	% des $x_i n_i$ cumulés croissants
0 - 10	5	16	16	20	80	80	4
10 - 20	15	30	46	57,5	450	530	25
20 - 40	30	18	64	80	540	1070	50
40 - 70	55	10	74	92,5	550	1620	76
70 - 100	85	6	80	100	510	2130	100
		80			2130		



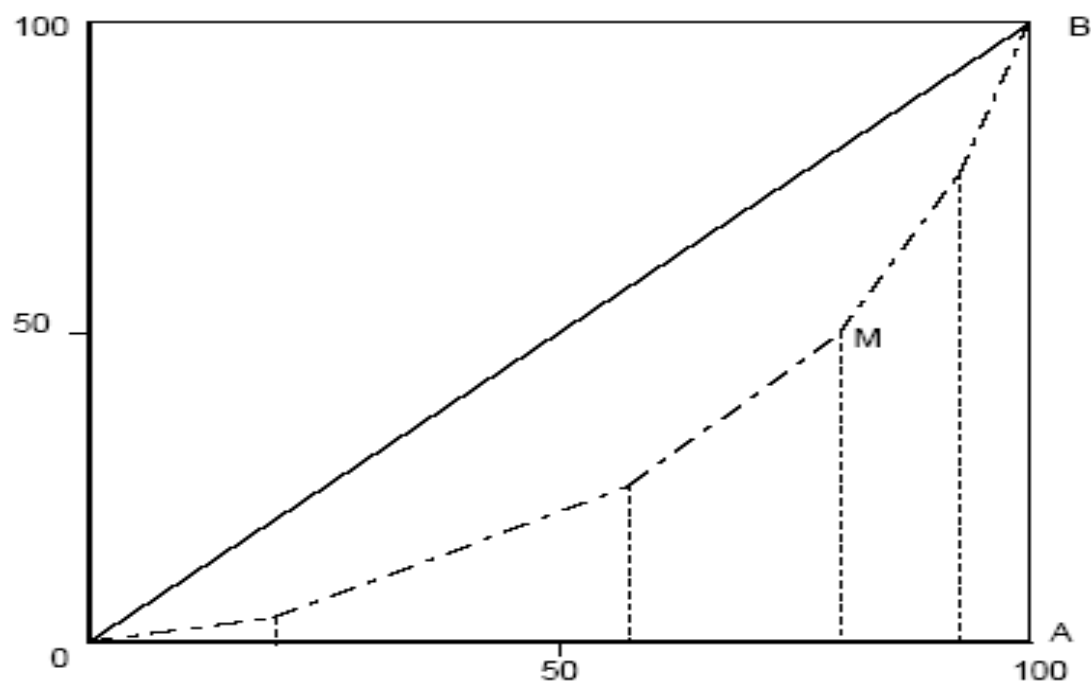
On peut construire un carré dont la base sera graduée de 0 à 100 et servira d'axe des abscisses et dont la hauteur sera graduée de 0 à 100 et servira d'axe des ordonnées.

En abscisse, on porte les effectifs cumulés, en pourcentage de l'effectif total.

En ordonnée, on porte les $x_i n_i$ cumulés, en pourcentage du total des $x_i n_i$.

On a les ordonnées suivantes :

% des Fréquence cumulée croissante	% des $x_i n_i$ cumulés croissants
20	4
57,5	25
80	50
92,5	76
100	100
Axe des abscices	Axe des ordonnés



La courbe O.M.B est dite courbe de concentration.

- Si toutes les parcelles avaient la même superficie, les pourcentages cumulés des nombres de terrains et les pourcentages cumulés des x_i n_i seraient égaux entre eux sur chaque ligne du tableau. La courbe de concentration se réduirait à la diagonale OB du carré. La concentration serait nulle.
- Plus la courbe de concentration est éloignée de la diagonale, plus la concentration est forte.

A) L'INDICE DE CONCENTRATION DE GINI :

On peut mesurer cette concentration au moyen de l'indice de concentration.

Il est égal au rapport entre :

- l'aire comprise entre la diagonale OB et la courbe OMB

et

- l'aire du demi carré = $(100 * 100) / 2 = 5000$

Cet indice varie entre 0 et 1

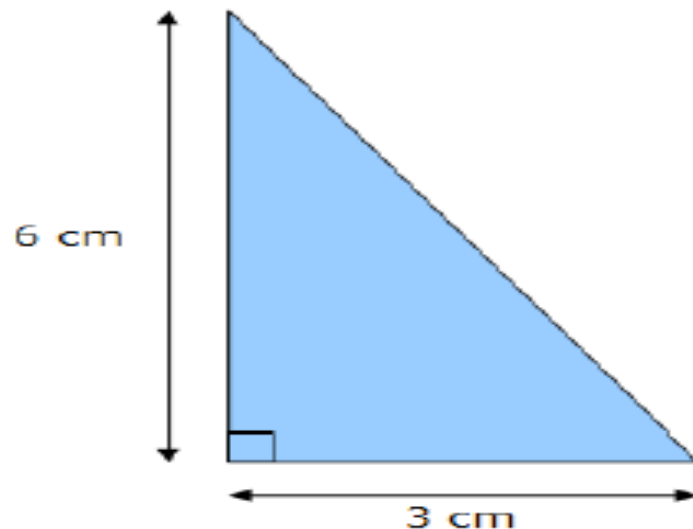
$$\begin{aligned}\text{Concentration } C &= \frac{\text{Aire de la concentration}}{\text{Aire du triangle OAB}} \\ &= \frac{\text{Aire de la concentration}}{5000}\end{aligned}$$

Aire de la concentration = aire du triangle OAB – aire de OABM

L'aire OABM est décomposée en un triangle rectangle et des trapèzes.

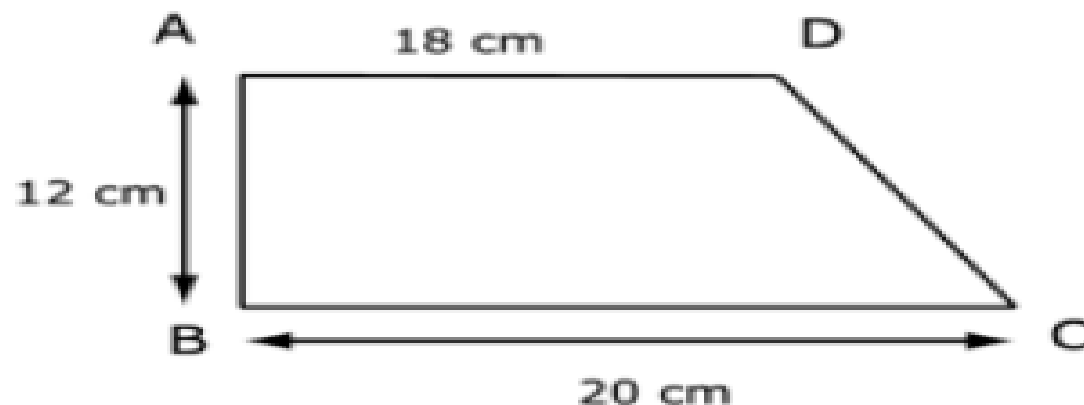
Sachant que :

$$\text{L'aire du triangle rectangle} = \frac{\text{Hauteur} * \text{Base}}{2}$$



$$\text{L'aire du triangle rectangle} = \frac{6 * 3}{2} = 9 \text{ cm}$$

Aire d'un Trapèze = $\frac{\text{Hauteur} * (\text{Petite Base} + \text{Grande Base})}{2}$



Aire du Trapèze : $\frac{12 * (18 + 20)}{2} = 228 \text{ cm}$

Maintenant on va calculer les surfaces d'OMB

- $S1 = \text{surface du triangle rectangle} = \frac{(4 * 20)}{2} = 40 \text{ cm}$
- $S2 = \text{surface du trapèze} = \frac{[(4 + 25)(57,5 - 20)]}{2} = 543,75 \text{ cm}$
- $S3 = \text{surface du trapèze} = \frac{[(25 + 50)(80 - 57,5)]}{2} = 843,75 \text{ cm}$
- $S4 = \text{surface du trapèze} = \frac{[(50 + 76)(92,5 - 80)]}{2} = 787,5 \text{ cm}$
- $S5 = \text{surface du trapèze} = \frac{[(76 + 100)(100 - 92,5)]}{2} = 660 \text{ cm}$

$$\begin{aligned}\text{Donc Aire OABM} &= 40 + 543,75 + 843,75 + 787,5 + 660 \\ &= 2\,875 \text{ cm}\end{aligned}$$

$$\begin{aligned}\text{Et Aire OMB} &= \text{Aire OAB} - \text{Aire OABM} \\ &= 5\,000 - 2\,875 = 2\,125\end{aligned}$$

$$\text{Indice de concentration} = \frac{2125}{5000} = 0,425 = 42,5\% : \text{c'est une concentration moyenne.}$$