Chapitre III : Les caractéristiques de dispersion

Les caractéristiques de tendance centrale ne sont pas toujours suffisantes pour caractériser une série statistique, car 2 séries peuvent avoir Mo= Me = $\bar{x}$ alors qu'elles sont distribuées de façon différente comme le montre l'exemple suivant :

Exemple : on a 2 séries relatives au nombre de pièces mauvaises produites par 2 machines A et B par semaine.

Machine A : 39, 40, 40, 41, 41, 41, 42, 42,43

Machine B : 1, 21, 21, 41, 41, 41, 61, 61,81

On constate que ces 2 séries ont : Me = Mo = $\bar{x}$ = 41 pièces mauvaises, cependant les 2 séries sont différentes par la dispersion de leurs valeurs.

Pour la série A : les valeurs se concentrent autour de la valeur centrale tandis que

Pour la série B : les valeurs sont plus dispersées par rapport à la valeur centrale.

Pour mesurer le degré de dispersion de ces valeurs par la valeur centrale on fait appel a quelques caractéristiques de dispersion telles : l'entendue (E) ; les quantiles ; la variance et l'écart type.

I - L'étendue (E) ou l'intervalle de variation :

L'étendue notée (E) ou l'intervalle de variation d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la variable statistique

$$\boxed{E = \text{Limite finale} - \text{limite initiale}}$$

Exemple précèdent :

Machine A : E = 43 – 39 = 4

Machine B : E = 81 – 1 = 80

Remarques : - l'étendue est facile à calculer et contient 100% des observations.

- cette caractéristique n'est pas très fiable car elle dépend des valeurs extrêmes du caractère, valeurs souvent aberrantes ou accidentelles.

II- Les quantiles :

Ce sont des paramètres qui nous renseignent sur la dispersion des valeurs de la variable par rapport à la médiane.

A- Les quartiles

1- Définition **:** les quartiles Q_i sont des valeurs de variables qui partagent la série en 4 parties égales ; on a 3 quartiles notés Q_1 , Q_2 , Q_3 .

* Le premier quartile Q_1 est la valeur du caractère tel que 25% des observations de la série ont une valeur inférieure à Q_1 et 75% des observations ont une valeur supérieure à Q_1.

* Q_2 c'est le $2^{éme}$ quartile qui correspond à la médiane Q_2 =Me.

* Le 3^{eme} quartile Q_3 : c'est la valeur du caractère tel que 75% des observations de la série ont une valeur inférieure à Q_3 et 25% des observations ont une valeur supérieure à Q_3.

$$\overbrace{0 \_\_\_\_ Q_1 \_\_\_\_ Q_2 \_\_\_\_ Q_3 \_\_\_\_}^{E} \_\_\_\_$$

25% 25% Me 25% 25% 100%

2- *L'intervalle interquartile* : c'est la différence entre Q_3 et Q_1 noté : Iq = $Q_3 - Q_1$
cet intervalle contient 50% des observations centrales, on peut le comparer à l'étendue pour mesurer son importance par rapport à la série.

3- Détermination des quartiles Q_1 et Q_3.

Comme la médiane Me, les quartiles se déterminent à partir des ni cum ↗ ou f_i cum ↗ et ce graphiquement ou par le calcul.

a/ calcul de Q_1, Q_3, Me.

Exemple : répartition de 100 *salariés d'une entreprise se selon le salaire mensuel en (1000 DH)*

Salaires	ni	ni cumulé croissant
8 – 10	22	22
10 – 12	28	50
10 – 14	30	80
14 – 16	15	95
16 - 18	5	100
$\sum$	100	-

*Q_1 : le rang de Q_1 : $\frac{\sum n_i}{4} = \frac{100}{4} = 25 \Rightarrow Q_1 \in [10 - 12[$. Calcul :

$$Q_1 = L_1 + \left(\frac{\sum n_i/4 - \sum_1 n_i}{\sum_2 n_i - \sum_1 n_i} \times (L_2 - L_1) \right) = 10 + \left(\frac{25 - 22}{50 - 22} \times 2 \right) = 10{,}21 \text{ soit}$$

10210 DH: interprétation: 25% des salariés touchent moins de 10210 DH
75% des salariés touchent plus de 10210 DH

*La médiane Me : le rang : $\sum n_i/2 = 100/2 = 50 \Rightarrow Me \in [10 - 12[$. Calcul de Me :
Me = 12 soit 12000 DH, car le rang apparaît dans le cumul croissant des n_i , la médiane correspond à la limite supérieure de la classe médiane (même constatation pour les autres quantiles).

*Le $3^{ème}$ quartiles Q_3 : le rang : $3\sum n_i = 100 \times 3 = 75 \Rightarrow Q_3 \in [12 - 14[$. Calcul :

$$Q_3 = L_1 + \left((L_2 - L_1) \times \frac{3\sum n_i/4 - \sum_1 n_i}{\sum_2 n_i - \sum_1 n_i} \right) = 12 + \left(2 \times \frac{75 - 50}{80 - 50} \right) = 13{,}667 \text{ soit } 13667 \text{ DH}$$

75% des salariés perçoivent un salaire < à 13667 DH et 25% des salariés ont un salaire ≥ à 13667 DH

4- L'intervalle interquartile $I_q = Q_3 - Q_1 = 3457$ DH. L'étendue des 50% des observations centrales est de 3457 DH.

B- les déciles (Di) et les centiles (Ci)

1- les déciles : on divise la population en 10 parties égales on obtient 9 déciles D_1 , D_2 ………D_9 .

* Le premier décile D_1 est la valeur du caractère telle que 10% des observations de la série ont une valeur inférieure à D_1 et 90% des observations ont une valeur supérieure à D_1 .

* Le 9^{eme} décile D_9 : c'est la valeur du caractère tel que 90% des observations de la série ont une valeur inférieure à D_9. et 10% des observations ont une valeur supérieure à D_9.

Remarque : l'intervalle interdécile D_9 - D_1 contient 80% des observations

détermination des déciles graphiquement ou par calcul. On utilise les ni cumulatif.

* Calcul de D_1 et D_9 :

Calcul des déciles D_1 et D_9 :

*D_1: le rang: $\frac{\sum n_i}{10} = \frac{100}{10} = 10 \Rightarrow D_1 \in [8 - 10[$: $D_1 = L_1 + \left((L_2 - L_1) \times \frac{(\sum n_i)/10 - \sum_1 n_i}{\sum_2 n_i - \sum_1 n_i} \right) = 8 + \left(2 \times \frac{10 - 0}{22 - 0} \right)$

$D_1 = 8,909$ soit 8909 DH.

*Calcul de D_9: le rang: $\frac{9 \sum n_i}{10} = \frac{100 \times 9}{10} = 90 \Rightarrow D_9 \in [14 - 16[$: $D_9 = L_1 + \left((L_2 - L_1) \times \frac{9(\sum n_i)/10 - \sum_1 n_i}{\sum_2 n_i - \sum_1 n_i} \right)$

$D_9 = 14 + \left(2 \times \frac{90 - 80}{95 - 80} \right) = 15,333$ soit 15333 DH

2- L'intervalle interdécile : $I_D = D_9 - D_1$; il contient 80% des observations centrales.

$I_D = 15333 - 8909 = 6424$ DH. 80% des observations centrales ont des salaires occupant une étendue de 6424 DH. Celle-ci représente $64,24\% = \frac{6,424 \times 100}{10}$ de l'étendue totale (E = 18 – 8 = 10).

2- Les centiles Ci

On divisent les observations d'une série statistique en 100 parties égales on obtient 99 centiles.

* Le premier centile C_1 est la valeur du caractère telle que 1% des observations de la série ont une valeur inférieure à C_1 et 99% des observations ont une valeur supérieure à C_1 .

* Le 99^{eme} centile C_{99} : c'est la valeur du caractère tel que 99% des observations de la série ont une valeur inférieure à C_{99}. et 1% des observations ont une valeur supérieure à C_{99}.
Remarque : l'intervalle intercentile C_{99} - C_1 contient 98% des observations

a- Détermination des centiles C_1 et C_{99}. On utilise les ni cumulatifs.

III - La variance V(x) et l'écart type (σx)

A- La variance V(x) :

1- Définition : la variance est la moyenne arithmétique des carrés des écarts de la variable x_i à sa moyenne arithmétique $\bar{x}$.

Elle se définit par la quantité suivante :

$$\left.\begin{array}{l} si.n_i = 1 : V(x) = \dfrac{\sum (xi - \bar{x})^2}{n} \\ si.n_i \neq 1 : V(x) = \dfrac{\sum n_i (xi - \bar{x})^2}{\sum n_i} \end{array}\right\} \text{formule de définition.}$$

- En développant la formule de définition on obtient une formule développée :

$$V(x) = \frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i} = \frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}^2 = \sum f_i x_i^2 - \bar{x}^2 = \frac{\sum f_i \% x_i^2}{100} - \bar{x}^2 = Q^2 - \bar{x}^2 .$$

2- l'écart type (σx) : l'écart type est la racine carrée de la variance V(x)

$$\sigma x = \sqrt{V(x)} = \left[\frac{\sum n_i (x_i - \bar{x})^2}{\sum n_i} \right]^{1/2}$$

L'écart type montre de combien en moyenne, la variable x_i s'écarte de sa moyenne $\bar{x}$.

3- Le coefficient de variation C.V.

Le C.V. est le rapport de l'écart type à la moyenne qui s'exprime en %

$$C.V. = \frac{\sigma_x}{\bar{x}} \times 100.$$

On utilise le CV pour apprécier l'importance de la dispersion d'une série ou pour comparer deux séries entre elles .

B/ Calcul de V(x), σx et C.V.

On reprend l'exemple précèdent : le salaire mensuel (10^3 DH)

Exercice d'application : exemple précédent : calculer V(x), δx et C.V.

Salaire 10^3 DH	n_i	Centre de classe x_i	$n_i x_i$	$(x_i - \bar{x})$	$n_i (x_i - \bar{x})^2$
8 – 10	22	(8+10)/2= 9	22 x 9 = 198	– 3,06	205,92
10 – 12	28	(10+12)/2=11	308	– 1,06	31,36
12 – 14	30	(12+14)/2=13	390	+ 0,94	26,40
14 – 16	15	(14+16)/2=15	225	+ 2,94	129,60
16 – 18	5	(16+18)/2=17	85	+ 4,94	122,00
Σ	100	—	1206	—	515,28

$\bar{x} = \dfrac{\sum n_i x_i}{\sum n_i} = \dfrac{1206}{100} = 12,06$

$V(x) = \dfrac{\sum n_i (x_i - \bar{x})^2}{\sum n_i}$

$= \dfrac{515,28}{100} - (12,06)^2$

$= 5,1528$

$\Rightarrow \delta x = \sqrt{5,1528} = 2,27$

soit 2270 DH

Le C.V. = $\dfrac{2,27}{12,06} \times 100 = 18,82\%$

Interprétation : * δx = 2270 DH : les valeurs x_i de la variable x s'écartent de leur moyenne $\bar{x}$ = 12060 DH, en moyenne de 2270 DH.

* le C.V. = 18,82% la dispersion entre les salaires est faible ; salaires homogènes.