Chapitre 5 : Statistique descriptive

I Autour de la médiane :

1. Définition : médiane

Dans une série ordonnée de nombres, la médiane est la valeur qui partage la série en 2 sous-séries de même effectif.

Exemple : Série d'effectif N impair

Exercice 1 : a) 1 – 2 – 5 – 5 – 10 – $\boxed{15 – 29 – 36 – 43}$ $\quad N = 9$

la série statistique compte autant de valeurs inférieures à 10 que de valeurs supérieures à 10.

b) Série d'effectif N pair

Exemple : $\boxed{3 – 8 – 9 – 12 – 13}$ 15 – 17 – 20 – 23 – 25 $\quad N = 10 \quad \frac{N}{2} = 5$

$$M_e = \frac{13+15}{2} = 14$$

la médiane dans ce cas n'est pas une valeur de la série.

c) 18 – 11 – 2 – 3 – 9 – 21 – 7 – 13

on commence par ordonner la série : 2 – 3 – 7 – 9 11 – 13 – 18 – 21 $\quad N = 8 \quad \frac{N}{2} = 4$

$$M_e = \frac{9+11}{2} = 10$$

2. Quartiles : définition

- On appelle premier quartile et on note Q_1 la valeur de la série telle que au moins 1/4, soit au moins 25% des valeurs de la série lui sont inférieures ou égales.
- On appelle troisième quartile et on note Q_3 ~~la~~ valeur de la série telle que au moins 3/4 soit au moins 75% des valeurs de la série lui sont inférieures ou égales.

- L'écart interquartiles d'une série statistique est la différence $Q_3 - Q_1$

Remarque : L'intervalle interquartile $[Q_1 ; Q_3]$ contient la moitié centrale des valeurs de la série. Plus l'écart interquartile est important, plus les valeurs de la série sont dispersées autour de la médiane M_e.

3- Diagramme en boîte (ou boîte à moustache)

On peut résumer graphiquement une série statistique par un diagramme en boîte (boxplot en anglais) sur lequel figurent les indicateurs Min, Q_1, Me, Q_3, Max

Exemple : construire le diagramme en boîte de la série suivante d'effectif total $N = 10$

6 ; 9 ; 54 ; 11 ; 248 ; 23 ; 143 ; 305 ; 65 ; 135

on commence par ordonner la série : 6 - 9 - 11 - 23 - 54 - 65 - 135 - 143 - 248 - 305

(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})

$Me = \frac{54+65}{2} = 59,5$

$\frac{N}{4} = \frac{10}{4} = 2,5$

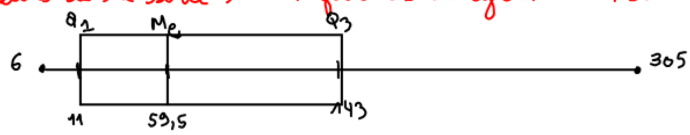
Q_1 est la valeur de la série telle qu'au moins 25% des valeurs de la série lui sont inférieures ou égales.

Au moins 2,5 valeurs, donc au moins 3 valeurs sinon on a moins de 25%. $Q_1 = x_3 = 11$

De même $\frac{3N}{4} = 7,5$ donc $Q_3 = x_8 = 143$.

Au moins 75% des valeurs de la série sont inférieures ou égales à 143.

Min = 6 Max = 305



4- Déciles : définition

On a défini les quartiles en prenant des quarts de l'effectif total de la série.

De la même façon on définit les déciles en prenant des dixièmes de l'effectif total.

1er décile, noté D1 : c'est la valeur de la série telle que au moins 10% des valeurs de la série lui sont inférieures ou égales.

9e décile, noté D9 : c'est la valeur de la série telle que au moins 90% des valeurs de la série lui sont inférieures ou égales

Exemple : On considère le tableau des salaires mensuels d'une entreprise

Salaire brut x_i (€)	1405	1480	1554	1870	2739	4215	Total
effectifs n_i	40	15	81	35	9	5	$N = 185$
ECC	40	55	136	171	180	185	

$\frac{N}{2} = 92,5$. $Me = x_{93} = 1554$

$x_1 \dots x_{92}$ x_{93} $x_{94} \dots x_{185}$

$\frac{N}{4} = \frac{185}{4} = 46,25$ $Q_1 = x_{47} = 1480$

$\frac{N}{10} = 18,5$ $D_1 = x_{19} = 1405$

$\frac{9N}{10} = 166,5$ donc $D_9 = x_{167} = 1870$; $\frac{3N}{4} = 138,75$ $Q_3 = x_{139} = 1870$

II Autour de la moyenne :

1. Moyenne : définition

Soit $(x_i ; n_i)$ avec $1 \leq i \leq p$ une série statistique quantitative dont les modalités $x_1 ; x_2 ; \dots ; x_p$ ont pour effectifs respectifs $n_1 ; n_2 ; \dots ; n_p$ et pour fréquences $f_1, f_2 ; \dots ; f_p$.

Son effectif total est $N = \sum_{i=1} n_i = n_1 + n_2 + \dots + n_p$

valeurs x_i	x_1	x_2	...	x_p	Total
effectifs n_i	n_1	n_2	...	n_p	N
fréquences f_i	f_1	f_2	...	f_p	1

La moyenne de la série statistique $(x_i ; n_i)_{1 \leq i \leq p}$ est notée $\bar{x}$ et on a :

$$\bar{x} = \frac{\sum_{i=1}^{p} n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n_1 + n_2 + \dots + n_p} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p = \sum_{i=1}^{p} f_i x_i$$

Exemple 1 :

En lançant un dé cubique 25 fois on a obtenu la distribution suivante :

1	2	3	4	5	6
2	5	4	5	6	3

La moyenne de cette série statistique est :

$$\bar{x} = \frac{1 \times 2 + 2 \times 5 + 3 \times 4 + 4 \times 5 + 5 \times 6 + 6 \times 3}{25}$$

$$\bar{x} = \frac{92}{25} = 3{,}68$$

Exemple 2 : données réparties en classe.

Le tableau suivant donne la répartition en % des auditeurs d'une radio FM suivant l'âge. Estimons l'âge moyen d'un auditeur de cette radio

Age	[8 ; 12[	[12 ; 16[	[16 ; 18[	[18 ; 22[	[22 ; 30[
fréquence	0,14	0,24	0,24	0,25	0,13

$\bar{x} = 10 \times 0{,}14 + 14 \times 0{,}24 + 17 \times 0{,}24 + 20 \times 0{,}25 + 26 \times 0{,}13$
$= 17{,}22$

Lorsque les données sont réparties en classe, les calculs s'effectuent avec les centres de classes.

Pour mesurer la dispersion des valeurs de la série $(x_i ; n_i)_{1 \leq i \leq p}$ autour d'un réel k, on peut former la moyenne des carrés des écarts à k des valeurs de la série.

$$f(k) = \frac{1}{N}\left[n_1(x_1 - k)^2 + n_2(x_2 - k)^2 + \dots + n_p(x_p - k)^2 \right]$$

$$f(k) = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - k)^2$$

On montre que f admet un minimum pour $k = \bar{x}$ et ce minimum est :

$$f(\bar{x}) = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - \bar{x})^2$$

2) Variance et écart type : définition

Soit la série statistique $(x_i ; n_i)_{1 \leq i \leq p}$ d'effectif total N et de moyenne $\bar{x}$

Le réel $V = \frac{1}{N}\left[n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2 \right]$ est appelé variance de la série statistique.

$\sigma = \sqrt{V}$ est appelé écart type de la série.

Remarque : Les indicateurs V et σ mesurent la dispersion des valeurs de la série autour de sa moyenne ; ils sont donc naturellement associés à $\bar{x}$.

Exemple : On reprend la distribution des 25 lancés de dé

x_i	1	2	3	4	5	6
n_i	2	5	4	5	6	3

$\bar{x} = 3{,}68$

$$V = \frac{1}{N}\left[n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + n_3(x_3 - \bar{x})^2 + \dots + n_6(x_6 - \bar{x})^2 \right]$$

$$V = \frac{1}{25}\left[2(1 - 3{,}68)^2 + 5(2 - 3{,}68)^2 + 4(3 - 3{,}68)^2 + 5(4 - 3{,}68)^2 + 6(5 - 3{,}68)^2 + 3(6 - 3{,}68)^2 \right]$$

$V \simeq 2{,}30 \qquad \sigma = \sqrt{V} \simeq 1{,}52$