Ch 5

Statistiques

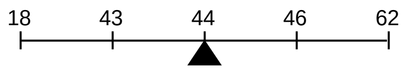
Durant ce chapitre nous allons nous intéresser à deux types d'indicateurs :

- les indicateurs de positions (médiane, moyenne) qui donne une position autour de laquelle les valeurs de la série
- les indicateurs de dispersion (étendue, écart interquartile, écart type) qui indique si les valeurs sont éloignées ou proches de la position.

I. Médiane

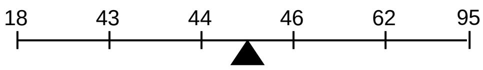
Définition
Soit une série statistique d'effectif total n, rangée par ordre croissant.
On appelle <u>médiane</u> la valeur "du milieu". On dit qu'elle partage la série en deux moitiés : il y a autant de valeurs en dessous qu'au dessus.
Pour déterminer son rang, il y a 2 cas :

- si n est impair : la médiane est la valeur de rang $\dfrac{n+1}{2}$



méd = 44

- si n est pair : nous prendrons la demi-somme des deux valeurs dont les rangs entourent le nombre $\dfrac{n+1}{2}$



méd = 45

Remarque :
Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. En revanche, on appellera <u>classe médiane</u>, la classe qui la contient (et permet donc d'en donner un encadrement).

Exemples
Données discrètes "en vrac"
10, 7, 12, 18, 16, 15, 5, 11, 11, 20, 15, 11, 18, 14
Ordonnons la série par ordre croissant : 5, 7, 10, 11, 11, 11, 12, 14, 15, 15, 16, 18, 18, 20
Il y a 14 termes or $\dfrac{14+1}{2} = 7,5$.

La médiane est donc la demi somme des 7$^{\text{ème}}$ et 8$^{\text{ème}}$ termes : méd $= \dfrac{12+14}{2} = 13$

Avec un tableau d'effectifs

valeurs	1	2	3	4	5	6
effectifs	6	11	25	19	15	5
effectifs cumulés	6	17	42	61	76	81

Attention, il faut bien interpréter cette dernière ligne : Les données qui valent 3 ont un rang compris entre 18 et 42 inclus

L'effectif total est de 81 or $\dfrac{81+1}{2} = 41$. La médiane est donc le 41$^{\text{ème}}$ terme : méd = 3

Avec des données réparties par classes

classe	[0 ; 2[	[2 ; 4[	[4 ; 6[	[6 ; 8]
fréquence	10%	38%	45%	7%
fréquence cumulée	10%	48%	93%	100%

48% des valeurs sont strictement inférieures à 4
Et 93% des valeurs sont strictement inférieures à 6
La classe médiane est donc la classe [4 ; 6[
On peut donc en déduire l'encadrement suivant $4 < \text{méd} < 6$

II. Moyenne

Exemple :
Soit la série statistique ci-contre :

valeurs	0	1	2	3	4
effectifs	1	2	1	4	2

La moyenne est : $\bar{x} = \dfrac{0+1+1+2+3+3+3+3+4+4}{1+2+1+4+2} = \dfrac{24}{10} = 2,4$

On préfèrera écrire : $\bar{x} = \dfrac{0+2\times1+2+4\times3+2\times4}{1+2+1+4+2} = \dfrac{24}{10} = 2,4$

Notation.
Le symbole Σ (sigma) signifie que l'on ajoute les éléments, par exemple

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

Formule pour le calcul de la moyenne
Soit la série statistique ci-contre :

valeurs	x_1	x_2	…	x_p
effectifs	n_1	n_2	…	n_p

La <u>moyenne</u> est : $\bar{x} = \frac{n_1x_1+n_2x_2+\ldots+n_{N-1}x_{N-1}+n_Nx_N}{n_1+n_2+\ldots+n_{N-1}+n_N} = \frac{\sum_{i=1}^{N} n_ix_i}{\sum_{i=1}^{N} n_i}$

Remarque :
Si les données ont été regroupées en classes, on ne peut calculer la valeur exacte de la moyenne. On peut toutefois en déterminer une bonne approximation en remplaçant chaque classe par son milieu.

Exemples :
a) Tableau de fréquences

valeurs	12	13	14	15	16
fréquences	0,05	0,17	0,43	0,30	0,05

$\bar{x}$ =

b) Données réparties en classes

Milieu de classe				
classes	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20]
effectifs	7	12	14	2

Remplaçons chaque classe par son milieu : $\frac{0+5}{2} = \ldots\ldots$

$\bar{x} = \frac{7\times\ldots\ldots+12\times\ldots\ldots+14\times\ldots\ldots+2\times\ldots\ldots}{} \approx$

III. Quartiles

Définitions :
Le premier quartile Q_1 est la plus petite des valeurs de la série telle qu'au moins 25% de la population ait sa valeur inférieur ou égale à Q_1.
Le troisième quartile Q_3 est la plus petite des valeurs de la série telle qu'au moins 75% de la population ait sa valeur inférieure ou égale à Q_3.

Méthode pour trouver Q_1 et Q_3.
Pour une population d'effectif n, $\frac{n}{4}$ (ou $\frac{3n}{4}$) si il est un entiers nous donne le rang de Q_1 (ou Q_3), dans le cas contraire on prendra la valeur de l'élément du rang immédiatement supérieur.

Exemple
Avec un tableau d'effectifs

valeurs	1	2	3	4	5	6
effectifs	6	11	25	19	15	5
effectifs cumulés	6	17	42	61	76	81

Ici n = 81 Donc $\frac{n}{4} = 20{,}25$ il va donc falloir que je prenne la valeur de l'individu de rang 21, donc $Q_1 = 3$
$\frac{3n}{4} = 60{,}75$ L'individu de rang 61 a pour rang 5 donc $Q_3 = 4$

Définitions :
L'intervalle interquartile est : $[Q_1; Q_3]$ 50% de la population a ses valeurs dans cet intervalle. Sa longueur sera nommée écart inter quartile est elle vaut $Q_3 - Q_1$, c'est un indicateur de dispersion bien plus fiable que l'étendue (rappel : étendue = Max – Min)

V. Ecart type et variance

Si on veut faire la moyenne des écarts entre les valeurs et leur moyenne on trouve 0, car les écarts positifs sont compensés par les écart positifs.

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

Pour régler le problème de signe on peut faire la moyenne des carrés des différences entre les valeurs et la moyenne et on obitent :

définitions

La variance est : $V = \frac{\sum_{i=1}^{N} n_i(x_i-\bar{x})^2}{\sum_{i=1}^{N} n_i}$

L'écart type est : $\sigma = \sqrt{V}$

Utilisation de la calculatrice

Si on a juste une série de valeur, on la rentre dans L_1 et on tapera : Var 1 Stat L_1

Si on a des effectifs associés à chaque valeur , alors après avoir rentré les valeurs dans L_1 et les effectifs dans L_2 on tapera : Var 1 Stat L_1, L_2

Methode de calcul

Si il n'y a que n valeurs sans effectif associé : $V = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{\text{n}}$ ou encore $V = \frac{{x_1}^2+{x_2}^2+\cdots+{x_n}^2}{\text{n}} - \bar{x}^2$

Si il y a p valeurs avec des effectifs associés : $V = \frac{n_1(x_1-\bar{x})^2+n_p(x_2-\bar{x})^2+\cdots+n_p\left(x_p-\bar{x}\right)^2}{n_1+n_2+\cdots+n_p}$ ou $V = \frac{n_1{x_1}^2+n_2{x_2}^2+\cdots+n_p{x_p}^2}{n_1+n_2+\cdots+n_p} - \bar{x}^2$

Si on a p intervalles de centres respectifs c_1, c_2, c_3 et c_p avec des effectifs associés :

$$V = \frac{n_1(c_1-\bar{x})^2+n_p(c_2-\bar{x})^2+\cdots+n_p\left(c_p-\bar{x}\right)^2}{n_1+n_2+\cdots+n_p} \text{ ou } V = \frac{n_1{c_1}^2+n_2c_2^2+\cdots+n_p{c_p}^2}{n_1+n_2+\cdots+n_p} - \bar{x}^2$$

Exemples

Déterminer les moyennes, variance et écart types des séries suivantes :

a) 4,7,9,12,13,13,19

b)

valeurs	1	2	3
effectifs	6	11	25

c)

Milieu de classe				
classes	[0 ; 10[	[10 ; 20[	[20 ; 30[	[30 ; 50[
effectifs	6	7	4	3