Statistiques descriptives
Variance et écart type

I) Rappel : la moyenne (caractéristique de position)

Définition

Soit la série statistique définie dans le tableau suivant :

Valeur	x_1	x_2	.....	x_p
Effectif	n_1	n_2	.....	n_p
Fréquences	f_1	f_2		f_p

Effectif total : $N = n_1 + n_2 + \ldots + n_p$ **et** $f_i = \frac{n_i}{N}$

La moyenne de cette série statistique est le réel, noté $\overline{x}$, tel que :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N}$$

ou en utilisant les fréquences :

$$\overline{x} = f_1x_1 + f_2x_2 + \cdots .. + f_px_p$$

Exemple 1: Soit la série statistique répertoriant la taille en mètres de 100 requins blancs

taille (en m)	1,5	2	2,5	3	3,5	4	4,5
Effectif	8	10	25	32	19	4	2

La taille moyenne est :

$$\overline{x} = \frac{1{,}5\times8 + 2\times10 + 2{,}5\times25 + 3\times32 + 3{,}5\times19 + 4\times4 + 4{,}5\times2}{100} = 2{,}82$$

Exemple 2 :

Un supermarché a relevé les dépenses (en €) de ses clients en 2 heures un jour donné, les résultats sont rassemblés dans le tableau suivant :

Dépenses (en €)	[0 ; 30 [	[30 ; 60 [	[60 ; 100 [	[100 ; 120 [
Milieu de classe	15	45	80	110
Effectif	12	25	42	67

Pour calculer la moyenne on détermine les milieux des classes de la distribution puis on effectue le calcul : $\overline{x} = \frac{15\times12 + 45\times25 + 80\times42 + 110\times67}{146} \approx 82{,}43$ €

(146 est l'effectif total)

Exemple 3 :

On étudie dans une maternité la taille de 50 nouveaux nés

Taille en cm	47	48	49	50	51	52
Effectif	5	8	12	15	9	1
Fréquence	0,1	0,16	0,24	0,3	0,18	0,02

$\overline{x} = 0{,}1 \times 47 + 0{,}16 \times 48 + 0{,}24 \times 49 + 0{,}3 \times 50 + 0{,}18 \times 51 + 0{,}02 \times 52 = 49{,}36$

II) Variance et écart type (caractéristique de dispersion)

Définition

Soit la série statistique définie dans le tableau suivant :

Valeur	x_1	x_2	.....	x_p
Effectif	n_1	n_2	.....	n_p
Fréquences	f_1	f_2		f_p

Effectif total : $N = n_1 + n_2 + \ldots + n_p$ **et** $f_i = \frac{n_i}{N}$

Soit $\overline{x}$ la moyenne de cette série .

Le réel V = $\frac{1}{N}[n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \cdots + n_i(x_i - \overline{x})^2 + \cdots + n_p(x_p - \overline{x})^2]$ **est appelé variance de cette série statistique.**

La racine carrée de la variance $\sigma = \sqrt{V}$ **est l'écart type de cette série.**

La **variance** et **l'écart type** permettent de mesurer la **« dispersion »** des valeurs de la série autour de la moyenne.
Si les valeurs de la série possèdent une unité, l'écart type s'exprime dans la même unité.

Autre formule pour calculer la variance :

V = $\frac{1}{N}[n_1{x_1}^2 + n_2{x_2}^2 + \cdots + n_i{x_i}^2 + \cdots + n_p{x_p}^2] - (\overline{x})^2$

Exemples :

Calculs de la variance et de l'écart type des séries précédentes

1°) Soit la série statistique répertoriant la taille en mètres de 100 requins blancs

taille (en m)	1,5	2	2,5	3	3,5	4	4,5
Effectif	8	10	25	32	19	4	2

La taille moyenne est :

$$\overline{x} = \frac{1{,}5\times8 + 2\times10 + 2{,}5\times25 + 3\times32 + 3{,}5\times19 + 4\times4+4{,}5\times2}{100} = 2{,}82$$

La variance $V = \dfrac{1{,}5^2\times8 + 2^2\times10 + 2{,}5^2\times25 + 3^2\times32 \ + \ 3{,}5^2\times19 \ + 4^2\times4 + 4{,}5^2\times2}{100} - 2{,}82^2$

(on utilise la deuxième formule)

$V = 8{,}395 - 7{,}9524 = 0{,}4426$ et $\sigma = \sqrt{0{,}4426} \approx 0{,}665$ m

2°) On étudie dans une maternité la taille de 50 nouveaux nés

Taille en cm	47	48	49	50	51	52
Effectif	5	8	12	15	9	1
Fréquence	0,1	0,16	0,24	0,3	0,18	0,02

$$\overline{x} = 0{,}1 \times 47 + 0{,}16 \times 48 + 0{,}24 \times 49 + 0{,}3 \times 50 + 0{,}18 \times 51 + 0{,}02 \times 52 = 49{,}36$$

• Disposition pratique de calcul de la variance et de l'écart type (avec la formule de la définition)

Taille en cm (xi)	Effectifs (ni)	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
47	5	$(47 - 49{,}36)^2 = 5{,}5696$	5*5,5696 =27,848
48	8	$(48 - 49{,}36)^2 = 1{,}8496$	8*1.8496 = 14,7968
49	12	$(49 - 49{,}36)^2 = 0{,}1296$	12*0.1296 = 1,5552
50	15	$(50 - 49{,}36)^2 = 0{,}4096$	15*0.4096 = 6,144
51	9	$(51 - 49{,}36)^2 = 2{,}6896$	9*2.6896 =24,2064
52	1	$(52 - 49{,}36)^2 = 6{,}9696$	1*6,9696 = 6,9696
Sommes	50		81,52

$V = \dfrac{81{,}52}{50} = 1{,}6304$ et $\sigma = \sqrt{V} \approx 1{,}277$ cm

• Disposition pratique de calcul de la variance et de l'écart type (avec la deuxième formule)

Taille en cm (xi)	Effectifs (ni)	$x_i^{\,2}$	$n_i x_i^{\,2}$
47	5	47^2 =2 209	5* 2 209 = 11 045
48	8	48^2 =2 304	8*2 304 = 18 432
49	12	49^2 =2 401	12*2 401 = 28 812
50	15	50^2 =2 500	15*2 500 = 37 500
51	9	51^2 =2 601	9*2 601 = 23 409
52	1	52^2 = 2 704	1*2 704 = 2 704
Sommes	50		121 902

$V = \frac{121\,902}{50} - 49{,}36^2 = 1{,}6304$ et $\sigma = \sqrt{V} \approx 1{,}277$ cm

Remarques

1) Vous devez savoir obtenir la moyenne et l'écart type d'une série statistique à l'aide de votre calculatrice ou d'un logiciel
2) L'écart type d'une série donne une idée de la dispersion des valeurs de la série autour de la moyenne. Il possède la même unité que les valeurs de la série.

II) Choix d'un couple (moyenne, écart type) ou (médiane, écart interquartile) pour l'étude d'une série

Vous avez étudié en classe de seconde la médiane et l'intervalle interquartile d'une série statistique.
Ce couple (médiane, intervalle interquartile) comme le couple (moyenne, écart type) permettent d'étudier une série statistique ou comparer plusieurs séries statistiques.

Le couple (moyenne, écart type) est néanmoins sensible aux valeurs extrêmes de la série alors que le couple (médiane, intervalle interquartile) ne l'est pas.

Seule une étude précise permet de déterminer le couple le plus approprié à chaque série.

Exemple :

Dans une entreprise qui doit fabriquer des supports en bois de 30 cm, on étudie les productions de 2 employés sur 100 supports fabriqués.
On obtient les résultats suivants :

Employé A :

Longueur en cm	26	27	28	29	30	31	32	33	34
Nombre de supports	2	7	13	18	20	17	16	5	2

Employé B :

Longueur en cm	26	27	28	29	30	31	32	33	85
Nombre de supports	2	2	3	16	48	21	5	2	1

Si on étudie les valeurs caractéristiques des deux séries on obtient :

	Médiane	Q1	Q3	Q3 – Q1	Moyenne	Ecart type
Employé A	30	29	31	2	29,99	1,797
Employé B	30	30	31	1	30.56	5.595

Si on observe le couple (médiane, intervalle interquartile) l'employé B semble être le plus précis, mais si on observe le couple (moyenne, écart type) c'est l'employé A qui est le meilleur.

On voit en observant les tableaux de résultats que l'employé B a raté une pièce si on retire cette donnée « aberrante » les résultats de la série de l'employé B deviennent :

Médiane	Q1	Q3	Q3-Q1	Moyenne	Ecart type
30	30	31	1	30,01	1,176

Ces résultats font de cet employé le plus régulier des deux. Reste à savoir pourquoi il a raté cette pièce ….. !