Statistiques descriptives
Variance et écart type

I) Rappel : la moyenne (caractéristique de position)

1) 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 : $\overline{x} = \dfrac{15\times12 + 45\times25 + 80\times42 + 110\times67}{146} \approx 82{,}43$ €

(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

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

2) Propriété 1

Si on ajoute le même nombre k à toutes les valeurs de la série statistique, la moyenne augmente de k

Exemple :

Dans l'exemple précédent on pourrait soustraire 50 à toutes les tailles on obtiendrait une nouvelle moyenne :

$\overline{y} = 0{,}1 \times (-3) + 0{,}16 \times (-2) + 0{,}24 \times (-1) + 0{,}3 \times 0 + 0{,}18 \times 1 + 0{,}02 \times 2 = -0{,}64$

et on retrouve $\overline{x}$ en rajoutant 50 à $\overline{y}$: $\overline{x} = -\ 0{,}64 + 50 = 49{,}36$

3) Propriété 2

Si on multiplie toutes les valeurs de la série statistique par un même nombre k, la moyenne est multipliée par k

Exemple :

En étudiant maintenant la masse de 50 nouveaux nés de la maternité on obtient :

Masse en kg	2,8	2,9	3	3,1	3,2
Effectif	14	10	18	7	1

On peut multiplier les masses par 10 on calcule ainsi une moyenne $\overline{y}$:

$$\overline{y} = \frac{28\times14+29\times10+30\times18+31\times7+32\times1}{50} = 29{,}42$$

et on retrouve la moyenne en divisant $\overline{y}$ par 10 : $\overline{x} = \dfrac{29{,}42}{10} = 2{,}942$

II) Variance et écart type (caractéristique de dispersion)

1) 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 : $N = n_1 + n_2 + \dots + n_p$ **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 :

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

Exemple :

Démonstration :

En reprenant la formule de définition :

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

En développant les carrés

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

En regroupant les termes

$$V = \frac{1}{N}[\, n_1{x_1}^2 + n_2{x_2}^2 + \cdots + n_i{x_i}^2 + \cdots . + n_p{x_p}^2] - 2\bar{x}\,[\frac{1}{N}(n_1x_1 + n_2x_2 + .. + n_ix_i + \cdots n_px_p)] + \bar{x}^2\,[\frac{1}{N}(n_1 + n_2 + \cdots + n_i + \cdots n_p)]$$

En simplifiant

V=$\frac{1}{N}\left[n_1{x_1}^2 + n_2{x_2}^2 + \cdots + n_i{x_i}^2 + \cdots . + n_p{x_p}^2\right] - 2\bar{x}\,\bar{x} + \bar{x}^2$

D'où le résultat **V =** $\frac{1}{N}\left[\boldsymbol{n_1{x_1}^2} + \boldsymbol{n_2{x_2}^2} + \cdots + \boldsymbol{n_i{x_i}^2} + \cdots . + \boldsymbol{n_p{x_p}^2}\right] - (\boldsymbol{\bar{x}})^2$

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 :

$$\bar{x} = \frac{1{,}5\times 8 + 2\times 10 + 2{,}5\times 25 + 3\times 32 + 3{,}5\times 19 + 4\times 4 + 4{,}5\times 2}{100} = 2{,}82$$

La variance $V = \dfrac{1{,}5^2\times 8 + 2^2\times 10 + 2{,}5^2\times 25 + 3^2\times 32 + 3{,}5^2\times 19 + 4^2\times 4 + 4{,}5^2\times 2}{100} - 2{,}82^2$

(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

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

• 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

$V = \frac{81{,}52}{50} = 1{,}6304$ 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 σ = $\sqrt{V}$ ≈ 1,277 cm

2) Propriété 1

Si on ajoute le même nombre k à toutes les valeurs de la série statistique, la variance et l'écart type ne changent pas

Exemple :

Dans l'exemple précédent on pourrait soustraire 50 à toutes les tailles on obtiendrait une nouvelle moyenne :

$\overline{y} = 0{,}1 \times (-3) + 0{,}16 \times (-2) + 0{,}24 \times (-1) + 0{,}3 \times 0 + 0{,}18 \times 1 + 0{,}02 \times 2 = -0{,}64$

On aurait ainsi V = $\frac{5(-3)^2+8(-2)^2+12(-1)^2+15(0)^2+9(1)^2\ 1(2)^2}{50}$ - $(-0{,}64)^2$ =

2,04 – 0,4096

V =1,6304

3) Propriété 2

Si on multiplie toutes les valeurs de la série statistique par un même nombre k, la variance est multipliée par k^2 et l'écart type est multiplié par $|k|$

Exemple :

En étudiant maintenant la masse de 50 nouveaux nés de la maternité on obtient :

Masse en kg	2,8	2,9	3	3,1	3,2
Effectif	14	10	18	7	1

On peut multiplier les masses par 10 on calcule ainsi une moyenne $\overline{y}$:

$$\overline{y} = \frac{28\times14+29\times10+30\times18+31\times7+32\times1}{50} = 29{,}42$$

et on retrouve la moyenne en divisant $\overline{y}$ par 10 : $\overline{x} = \frac{29{,}42}{10} = 2{,}942$

Pour la variance le calcul avec les masses données par le tableau donne

$$V = \frac{2{,}8^2\text{x}14+2{,}9^2\text{x}10+3^2\text{x}18+ 3{,}1^2\text{x}7+3{,}2^2\text{x}1}{50} - 2{,}942^2 = 0{,}012036$$

et $\sigma = \sqrt{0.012036} \approx 0{,}11$

En multipliant les masses par 10 on trouve :

$$V_1 = \frac{28^2\text{x}14+29^2\text{x}10+30^2\text{x}18+ 31^2\text{x}7+32^2\text{x}1}{50} - 29{,}42^2 = 1{,}2036 \text{ soit } 100 \times V$$

et bien sûr $\sigma_1 = \sqrt{V_1} \approx 1{,}1$ soit $10 \times \sigma$