Chapitre 2 : Séries statistiques à une variable

On appelle **série statistique** la suite des valeurs prises par une variable X surles individus. Les valeurs de la variable X sont notées $x_1, \ldots, x_i, \ldots, x_k$.

A.2.1Effectif, Fréquence, Pourcentage

-L'Effectif total noté **N**, est le nombre d'individus qui composent la population, on le note cardΩ=N.

-L'Effectif d'une valeur $\mathbf{x_i}$ noté $\mathbf{n_i}$ est le nombre d'individus associé à cette valeur.

-La Fréquence noté f_i **,** est donnée par $f_i = \frac{ni}{N}$**.** f_i peut s'exprimer en pourcentage**.**

Propriétés

$\sum_{i=1}^{k} n_i = N$, $\sum_{i=1}^{k} f_i = 1$, $0 \leq f_i \leq 1$.

l'effectif cumulé : Quand les valeurs du caractères sont ordonnées, on peut cumuler les effectifs de façon croissante ou décroissante.

● **L'effectif cumulé croissant (ECC)** d'une valeur (ou d'une classe) est la somme des effectifs de cette valeur (ou de cette classe) et des effectifs (ou des classes) précédentes

L'effectif cumulé décroissant (ECD) d'une valeur (ou d'une classe) est la somme des effectifs de cette valeur (ou de cette classe) et des effectifs (ou des classes) suivantes.

La fréquence cumulée croissante(FCD)d'une valeur (ou d'une classe) est la somme des **fréquences** de cette valeur (ou de cette classe) et des **fréquences** (ou des classes) précédentes

La Fréquence cumulée décroissante(FCD)d'une valeur (ou d'une classe) est la somme des **fréquences** de cette valeur (ou de cette classe) et des **fréquences** (ou des classes) suivantes.

A.2.2 Représentation des données

1-Tableau statistique

Un tableau statistique est un moyen d'organiser, classifier et rangerpar ordre croissant (ou décroissant) les données brutes de la série statistique, pour les biens représenter.

Exemple9 : Dans un centre de vacances, l'âge des personnes se répartit de la façon suivante :

Âge	[13 ; 14[	[14 ; 15 [	[15 ; 16[	[16 ; 17[
Effectif	8	12	16	14

Âge	[13 ; 14[	[14 ; 15[	[15 ; 16[	[16 ; 17[	Total
c_i					
Effectif	8	12	16	14	
ECC					
ECD					
Fréquence					
FCC					
FCD					

Le centre d'une classe [a; b[est $c = \frac{a+b}{2}$

2-Représentations graphiques

Les représentations graphiques permettent de résumer visuellement des séries statistiques.

Le choix du type de graphe dépend de la nature des variables.

a) **Cas d'une variable statistique qualitative**

- **Diagramme circulaire** (ou camembert)

Le principe du graphe consiste à diviser un cercle ou un disque en secteurs. Chaque secteur représente une modalité sa surface est proportionnelle à la fréquence de cette modalité. Pour une modalité donnée x_i, l'effectif n_i, l'angle au centre αi correspondant est N →360

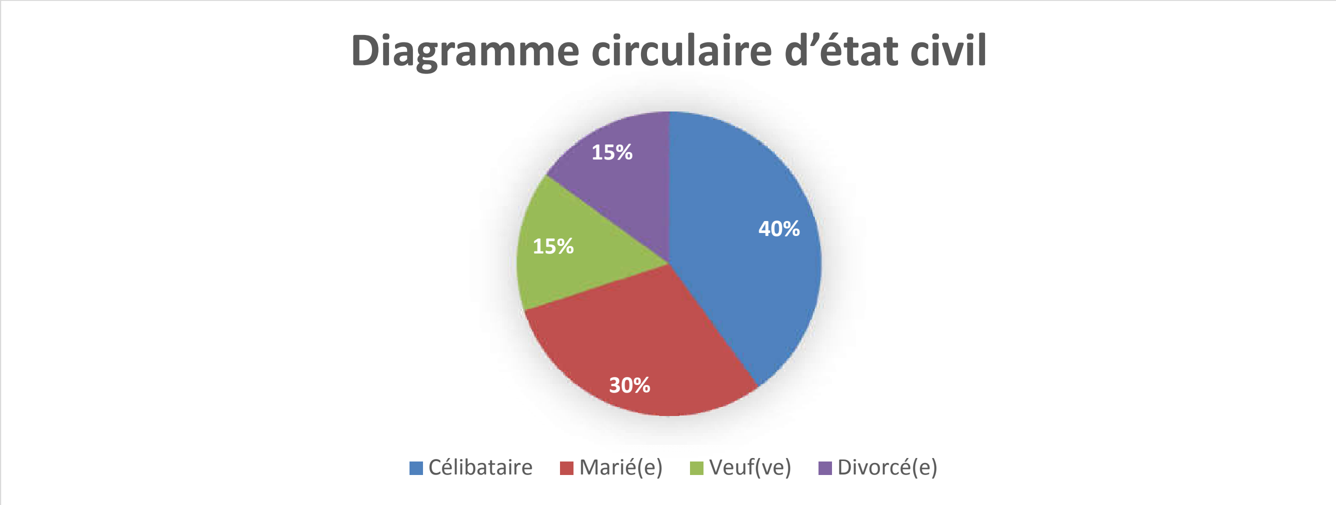
$n_i \rightarrow \alpha_i$, alors $\alpha i = \frac{ni}{N} * 360.$

N.B

$$\alpha i = fi * 360.$$

Exemple10 soitle tableau d'état civil suivant :

Les modalities x_i	Célibataire	Marié(e)	veuf(ve)	Divorcé(e)	total
Effectif n_i	8	6	3	3	20
Fréquence	0.4	0.3	0.15	0.15	1
Angle	144	108	54	54	360



- **Diagramme à bande** (**tuyaux d’orgue**)

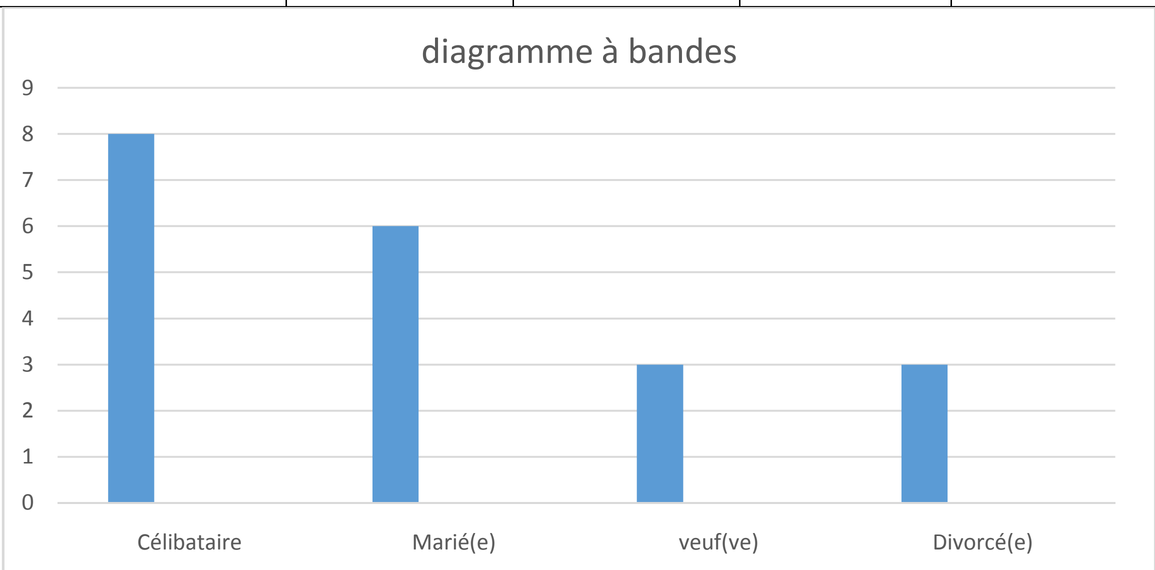
Un diagramme **à bande** est une représentation graphique de données à l'aide de rectangles de même largeur.

Les modalités sont représentées sur l'axe horizontal, les effectifs sur l'axe vertical. À chaque modalité correspond une barre. Les hauteurs des barres sont proportionnelles aux effectifs représentésla modalité.

N.B Le diagramme à bandes est utile aussi pour représenter des données discontinues.

Exemple10 soitle tableau d’état civil suivant :

Les modalités xi	Célibataire	Marié(e)	veuf(ve)	Divorcé(e)	total
Effectifn_i	8	6	3	3	20
Fréquence f_i	0.4	0.3	0.15	0.15	1



b) Cas d'une variable statistique quantitative

Pour les variables quantitatives, il existe deux types de représentation graphique qui sont :

1) <u>Cas d'une variable quantitative discrète :</u>

- **Diagramme à bâton (en bâton)(barres)**

Ce diagramme comporte deux axes, un axe horizontal qui représente les valeurs de la variable, et un axe vertical qui représente les Effectif. ou les fréquences, à chaque valeur on associe un segment (bâton) dont sa hauteur est proportionnelle à Effectif ou à la fréquence de cette modalité.

Exemple11

le tableau statistique suivant donne le nombre d'enfants à charge par famille dans un village

x_i	0	1	2	3	4	5	6
n_i	18	32	66	41	32	9	2

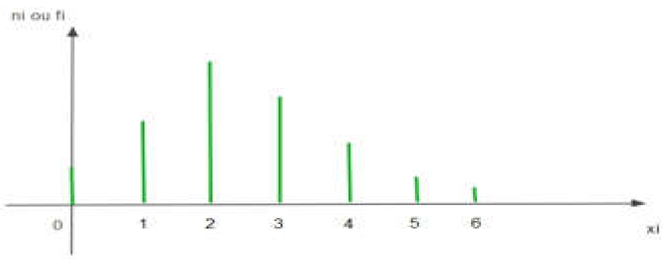


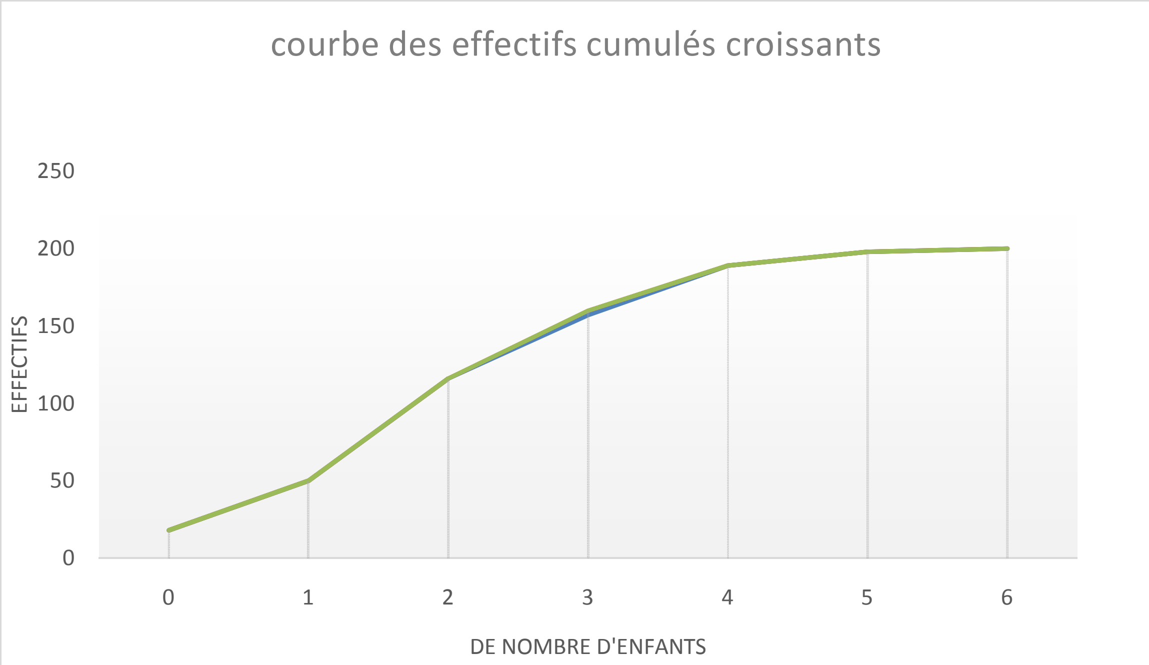
FIGURE 2.5: Diagramme à bâtons

- **Courbe des effectifs cumulés croissants**

  Pour tracer la courbe des effectifs cumulés croissants (ou décroissants):
  -On complète le tableau en calculant les effectifs cumulés croissants (ou décroissants).
  -On trace dans un repère le point dont l'abscisse est la modalité etl'ordonnée l'effectif cumulé qui lui correspond..
  On trace les segments qui relient chaque point du plan
  Exemple12

On complète le tableau statistique précédent :

x_i	0	1	2	3	4	5	6	total
n_i	18	32	66	41	32	9	2	200
ECC	18	50	116	157	189	198	200	



NB:

Cas d'une variable quantitative continue

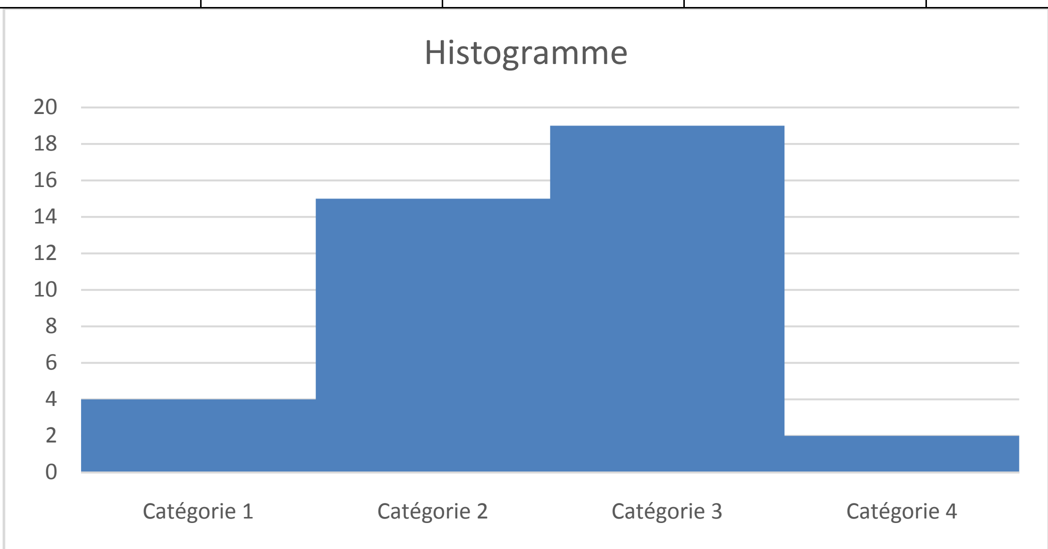
- **Histogrammes**

Un histogramme est constitué de rectangles juxtaposés ; la largeur dechaque rectangle correspond à l'intervalle de la classe correspondante ; sa hauteur est telle que l'aire du rectangle est proportionnelle à l'effectif de laclasse.

Exemple13

Une étude sur l'âge des femmes mariées d'un quartier d'une grande ville a fourni les informations suivantes

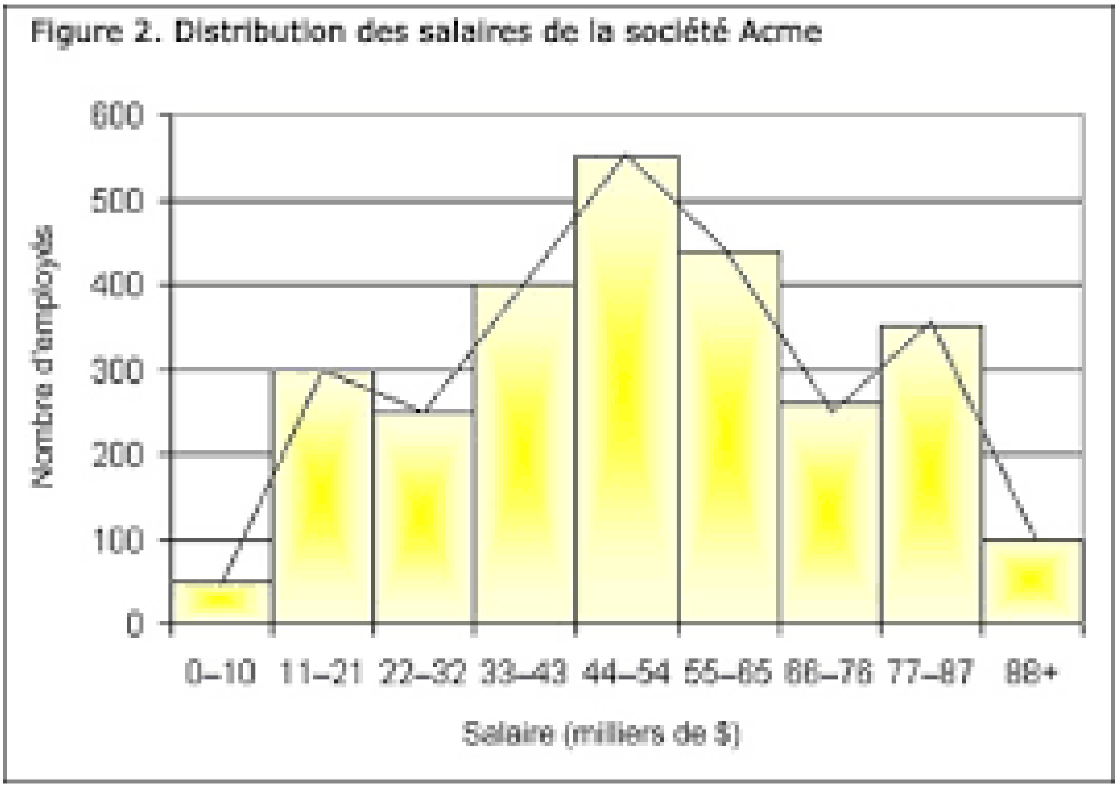
Age	*[18; 28[*	*[28; 38[*	*[38; 48[*	*[48; 58[*
Effectif	4	15	19	2



- **polygone des fréquences**

Le polygone des fréquences est représenté enjoignant les milieux des cotés supérieurs des rectangles dans un histogramme. C'est une ligne brisée dont les extrémités rejoignent l' axe des abscisses.

Exemple14



Paramètres de position (caractéristique de tendance centrale)

Les indicateurs statistiques de tendance centrale (dits aussi de position) considérés fréquemment sont la moyenne, la médiane et le mode

1. Le mode

1.1 Variable quantitative discrète : Notion de Mode

Définition 1 : Soit X une variable quantitative discrète, on appelle **mode** la valeur du caractère qui possède le plus grand effectif.

Exemple 1 : Dans le tableau suivant, représentant le nombre d'enfants par famille, le mode est '0 enfant'

Nombre d'enfants	0	1	2	3	4	5	6	7
Effectif (= nombre de foyers)	290	170	155	95	43	27	20	10

1.2 Variable quantitative continue : Notion de Classe modale

Définition 2: Soit X une variable quantitative continue, on appelle **classe modale** la classe du caractère qui possède le plus grand effectif.

Exemple 2 : Dans le tableau suivant, représentant les notes d'une classe, la classe modale est la classe [8 ; 12[

Notes	[0 ; 5[	[5 ; 8[	[8 ; 12[	[12 ; 15[	[15 ; 20]
Effectif	10	8	12	11	9

2. La moyenne

a) Définition

Soit la série statistique définie dans le tableau suivant :

Valeur	x_1	x_2	.....	x_p
Effectif	n_1	n_2	.....	n_p

Effectif total : N = $n_1 + n_2 + \cdots + n_p$

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

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_p x_p}{N}$$

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 :

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

- - -

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 : $\overline{x} = \dfrac{15\text{x}12 + 45\text{x}25 + 80\text{x}42 + 110\text{x}67}{146} \approx 82{,}43$ €

(146 est l'effectif total)

b) Propriété 1

On peut calculer la **moyenne** $\overline{x}$ à partir de la distribution des fréquences :

Valeur	x_1	x_2	.....	x_p
Fréquence	f_1	f_2	.....	f_p

$$\overline{x} = f_1\, x_1 + f_2\, x_2 + \ldots + f_p\, x_p$$

Exemple :

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

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

c) Propriété 2

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 :

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

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

d) Propriété 3

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 $\overline{y}$:

$$\overline{y} = \frac{28x14+29x10+30x18+31x7+32x1}{50} = 29{,}42$$

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

3. La médiane

a) Définition

La liste des N données est rangée par ordre croissant

- **Si N est impair (N = 2*n* + 1) la médiane est la donnée de rang *n* + 1**
- **Si N est pair (N = 2*n*) la médiane est la demi somme des données de rang *n* et de rang *n* + 1**

Exemple 1 :

Un boulanger teste les masses (en grammes) de 30 baguettes qu'il vient de fabriquer, il obtient les résultats suivants :

235	235	237	238	238	239	239	239	240	241
241	243	245	247	247	249	250	205	250	250
250	251	251	253	253	255	255	255	257	260

Comme l'effectif total N = 30 est pair la médiane est la demi somme de la donnée de rang 15 et la donnée de rang 16 soit : $\dfrac{247 + 249}{2} = 248$

Exemple 2 :

Le tableau ci-dessous indique la durée (en minutes) de connexion internet par jour de 43 familles interrogées

Durée en minutes	40	60	80	120	180	200	240	300
Effectif	2	9	11	7	5	2	4	3

Comme l'effectif total N = 43 = 2 x 21 + 1 est impair la médiane est la donnée de rang 22 soit 80 minutes

b) Propriétés

- Si on ajoute le même nombre k à toutes les valeurs de la série statistique, la médiane augmente de k
- Si on multiplie toutes les valeurs de la série statistique par un même nombre k, la médiane est multipliée par k

Paramètres de dispersion (variabilité)

Les indicateurs statistiques de dispersion usuels sont l'étendue, la variance et l'écart type.

1. L'étendue

La différence entre la plus grande valeur et la plus petite valeur du caractère, donnée par la quantité

$$e = x_{\max} - x_{\min},$$

s'appelle l'étendue de la V.S X. Le calcul de l'étendue est très simple. Il donne une première idée de la dispersion des observations.

2. La variance

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 :

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

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