Chapitre 7
Statistiques

I. Vocabulaire

Définitions :

L'ensemble sur lequel porte l'étude d'une **série statistique** s'appelle la **population**.

Un élément de la population est un **individu**.

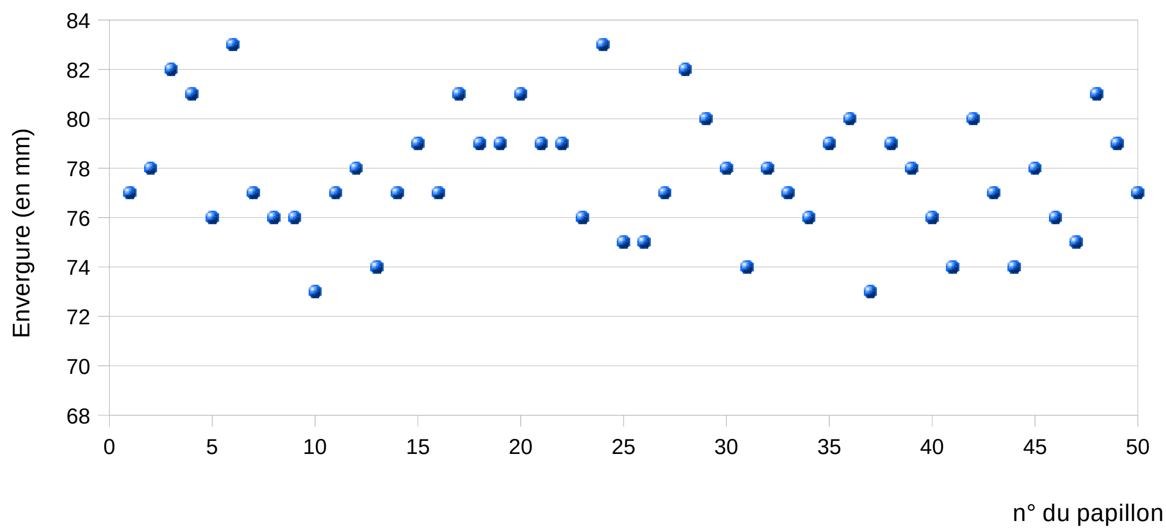
L'objet étudié s'appelle le **caractère** de la série.

- Si le caractère prend des valeurs **numériques**, on dit qu'il est **quantitatif**.
  Sinon, il est **qualitatif**.
- Un caractère quantitatif peut être **discret** ou **continu** :
  - discret s'il prend des valeurs isolées.
  - continu s'il peut prendre toute valeur dans un intervalle appelé aussi **classe**.
- La **fréquence** f d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total.

Exemple :

On a mesuré l'envergure (en *mm*) de 50 individus femelles d'une espèce de papillons et noté les valeurs obtenues :

77 ; 78 ; 82 ; 81 ; 76 ; 83 ; 77 ; 76 ; 76 ; 73 ; 77 ; 78 ; 74 ; 77 ; 79 ; 77 ; 81 ; 79 ; 79 ; 81 ; 79 ; 79 ; 76 ; 83 ; 75 ; 75 ; 77 ; 82 ; 80 ; 78 ; 74 ; 78 ; 77 ; 76 ; 79 ; 80 ; 73 ; 79 ; 78 ; 76 ; 74 ; 80 ; 77 ; 74 ; 78 ; 76 ; 75 ; 81 ; 79 ; 77

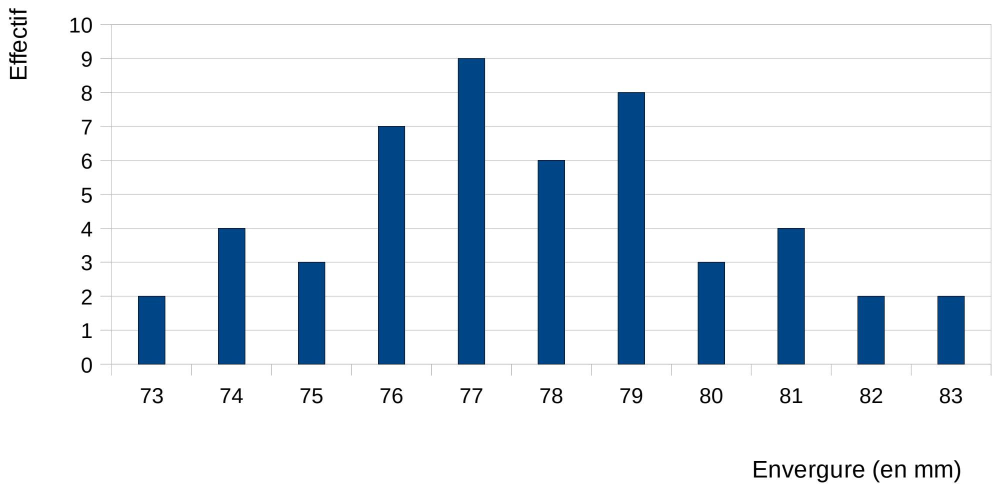


On obtient ainsi une **série brute**. On peut représenter la série par un **nuage de points**.

On a organisé la série précédente :

Envergures (en mm)	73	74	75	76	77	78	79	80	81	82	83
Effectifs	2	4	3	7	9	6	8	3	4	2	2

On peut choisir de représenter la série par un **diagramme en barre**.



Pour la **série dépouillée** précédente, on a :

Envergures (en mm)	73	74	75	76	77	78	79	80	81	82	83
Fréquences	0,04	0,08	0,06	0,14	0,18	0,12	0,16	0,06	0,08	0,04	0,04

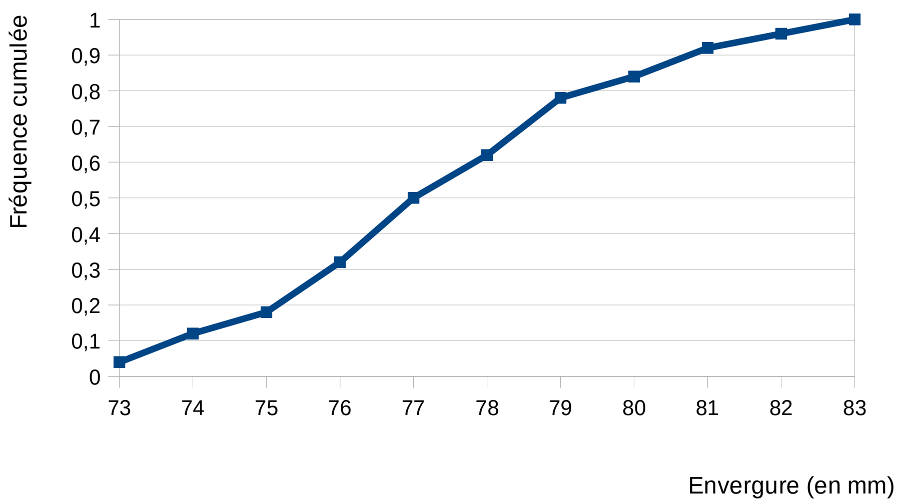
Dans l’exemple précédent, la **fréquence** de 73 est 0,04 : la fréquence de 74 est 0,08.

Pour 74, la **fréquence cumulée croissante** est 0,04 + 0,08 = 0,12 ;

pour 75, la fréquence cumulée croissante est 0,12 + 0,06 = 0,18 ; …

Envergures (en mm)	73	74	75	76	77	78	79	80	81	82	83
Fréquences	0,04	0,08	0,06	0,14	0,18	0,12	0,16	0,06	0,08	0,04	0,04
Fréquences cumulées croissantes	0,04	0,12	0,18	0,32	0,5	0,62	0,78	0,84	0,92	0,96	1

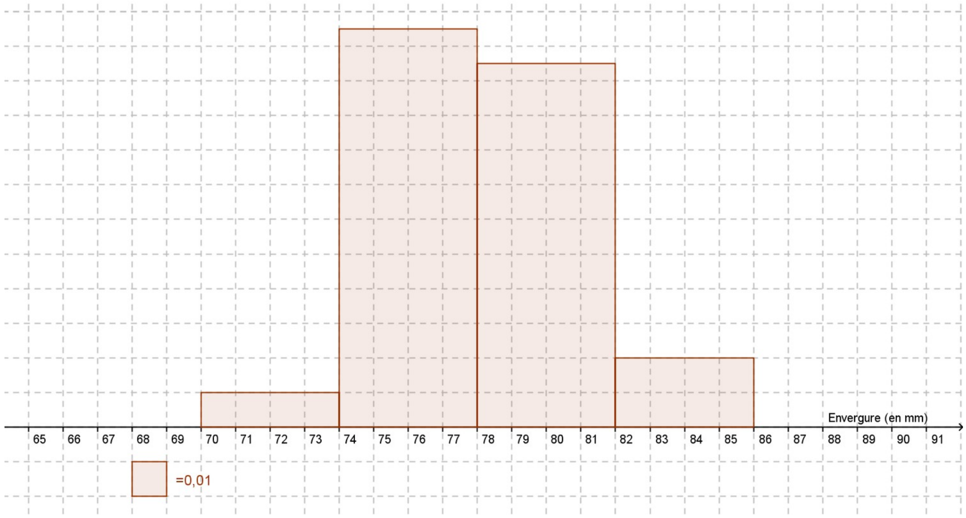
On peut alors représenter la série par une **courbe de fréquences cumulées**.



Dans le tableau suivant, la série est regroupée en **classes d'amplitude** 4.

Classes	[70 ; 74[	[74 ; 78[	[78 ; 82[	[82 ; 86[
Effectifs	2	23	21	4
Fréquences	0,04	0,46	0,42	0,08

On construit l'histogramme correspondant.



II. Médiane et écart interquartile

1) Médiane

Définition :

La **médiane** d'une série de N valeurs **rangées par ordre croissant** est le nombre Me, tel qu'**au moins 50 %** des valeurs de la série lui sont inférieures ou égales et au moins 50 % lui sont supérieures ou égales.

- Si N est **impair**, Me est la valeur centrale.
- Si N est **pair**, Me est la demi-somme des deux valeurs centrales.

2) Quartiles

Définition :

Les valeurs d'une série statistique étant **rangées par ordre croissant :**

- le **premier quartile** est la plus petite valeur Q_1 de la série telle qu'**au moins 25 %** des valeurs de la série sont inférieures ou égales à Q_1.
- le **troisième quartile** est la plus petite valeur Q_3 de la série telle qu'**au moins 75 %** des valeurs de la série sont inférieures ou égales à Q_3.

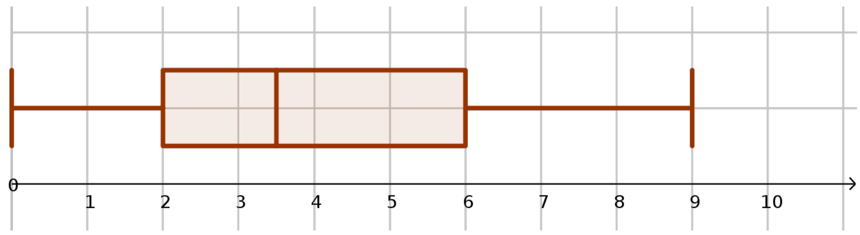
Exemple :

Soit la série ordonnée suivante :

0 ; 0 ; 0 ; 1 ; 2 ; 2 ; 2 ; 3 ; 3 ; 3 ; 3 ; 4 ; 4 ; 4 ; 4 ; 5 ; 6 ; 6 ; 6 ; 6 ; 6 ; 9

- L'effectif total est égal à 22 et est pair donc la médiane est la moyenne de la 11^e et la 12^e valeur de la série. Me = $\frac{3+4}{2}$ = 3,5.
- $22 \times 0,25 = 5,5$. Q_1 est la 6^e valeur de la série donc $Q_1 = 2$.
- $22 \times 0,75 = 16,5$. Q_3 est la 17^e valeur de la série donc $Q_3 = 6$.

On obtient le **diagramme en boîte** suivant :



3) Écart interquartile

Définitions :

- On appelle **écart interquartile** le nombre $Q_3 - Q_1$.
- L'**intervalle interquartile** est l'intervalle $[Q_1 ; Q_3]$.

Remarque :

Le couple (Me, $Q_3 - Q_1$) permet de résumer une série.

Il a l'avantage de ne pas être sensible aux valeurs extrêmes.

III. Moyenne et écart-type

On considère la série statistique donnée par le tableau ci-dessous.

N = n_1, n_2, … , n_p est l'effectif total.

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p
Fréquence	f_1	f_2	…	f_p

1) Moyenne pondérée

Définition :

La **moyenne pondérée** de cette série, notée $\overline{x}$, est donnée par la formule suivante :

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

Propriété :

Pour déterminer la **moyenne** à partir des fréquences, on utilise la formule :

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

Exemple :

On donne la série suivante :

Valeur	5	8	10
Effectif	1	2	2

La moyenne $\overline{x}$ est égale à $\overline{x}=\frac{5\times1+8\times2+10\times2}{1+2+2}$ = 8,2.

2) Linéarité de la moyenne

Propriété :

Lorsque toutes les valeurs de la série sont transformées par une fonction affine $x \mapsto mx + p$, la moyenne de la nouvelle série est $m\overline{x}+p$.

Exemple :

On reprend la série précédente et on multiplie toutes les valeurs de la série par 2 et on ajoute 1.

La série devient :

Valeur	11	17	21
Effectif	1	2	2

La nouvelle moyenne $\overline{x'}$ est égale à $\overline{x'}$ = $2\times\overline{x}+1$ = 17,4.

3) Écart-type

Définitions :

Soit la série de p valeurs x_1, x_2, … , x_p de moyenne $\bar{x}$.

- La **variance** V de la série est la moyenne des carrés des écarts des valeurs avec la moyenne :

$$V = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+...+(x_p-\bar{x})^2}{N} .$$

- L'**écart-type** σ de la série est la racine carrée de la variance $\sigma = \sqrt{V}$.

Exemple :

Soit la série 5 ; 8 ; 8 ; 10; 10 de moyenne 8,2.

La variance V est égale à $V = \frac{(5-8,2)^2+(8-8,2)^2+(8-8,2)^2+(10-8,2)^2+(10-8,2)^2}{N}$.

Soit $V = \frac{(5-8,2)^2+2\times(8-8,2)^2+2\times(10-8,2)^2}{N}$ = 3,36.

L'écart-type est égal à $\sigma = \sqrt{V}=\sqrt{3,36} \simeq 1,83$.

Remarque :

Le couple ($\bar{x}$, σ) permet de résumer une série.

Il a l'avantage d'utiliser toutes les valeurs de la série. En ce sens, il est représentatif de la série.

Il a l'inconvénient d'être sensible aux valeurs extrêmes de la série.