

# Chapitre 4

## Statistiques

### I. Vocabulaire

**Population :** Une population est un ensemble de personnes ou d'objets, appelés **individus**, définis par une propriété commune.

Exemple : les habitants d'un pays, les automobiles fabriquées en 2010,...

**Caractère :** Pour une population choisie, on peut étudier un caractère de ses individus.

Exemple : on peut étudier le caractère « taille » des élèves d'un lycée.

**Caractère quantitatif :** Un caractère est dit quantitatif lorsqu'il est possible de le mesurer en associant un nombre à chaque individu. Un caractère quantitatif est aussi appelé **variable**.

Exemple : l'âge, la taille, le nombre de frères et sœurs,...

- Un caractère quantitatif est dit **continu** lorsque les nombres qui le mesurent peuvent prendre, à priori, toutes les valeurs d'un intervalle.  
Exemple : le poids, la taille, la durée de vie d'un moteur,...
- Il est **discret** dans le cas contraire.  
Exemple : l'année de naissance, le nombre d'enfants par famille,...

**Caractère qualitatif :** On appelle ainsi tout caractère non quantitatif.

Exemple : la couleur des yeux.

### II. Présentation d'une série

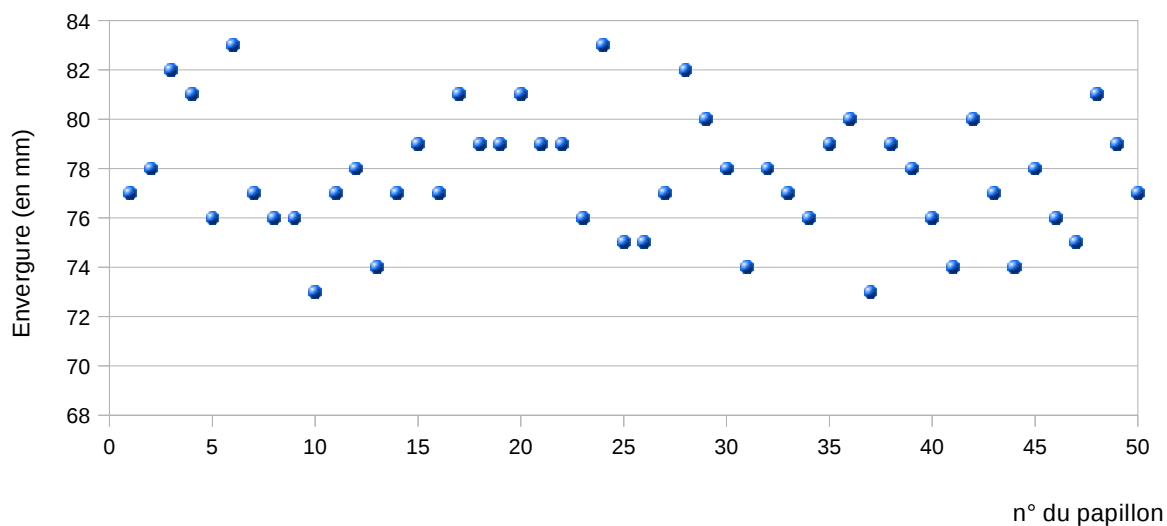
#### 1) Série statistique

#### Série brute

**Définition :** Une série quantitative **brute** est constituée de la liste de toutes les valeurs  $v_1, v_2, \dots, v_n$ . Le nombre de valeurs, ou **effectif**, de la série est  $n$ .

**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**.

## Série dépouillée

**Définition :** Une série est **dépouillée** quand on connaît la liste des valeurs prises  $x_1, x_2, \dots, x_r$  et l'effectif de chacune d'elles.

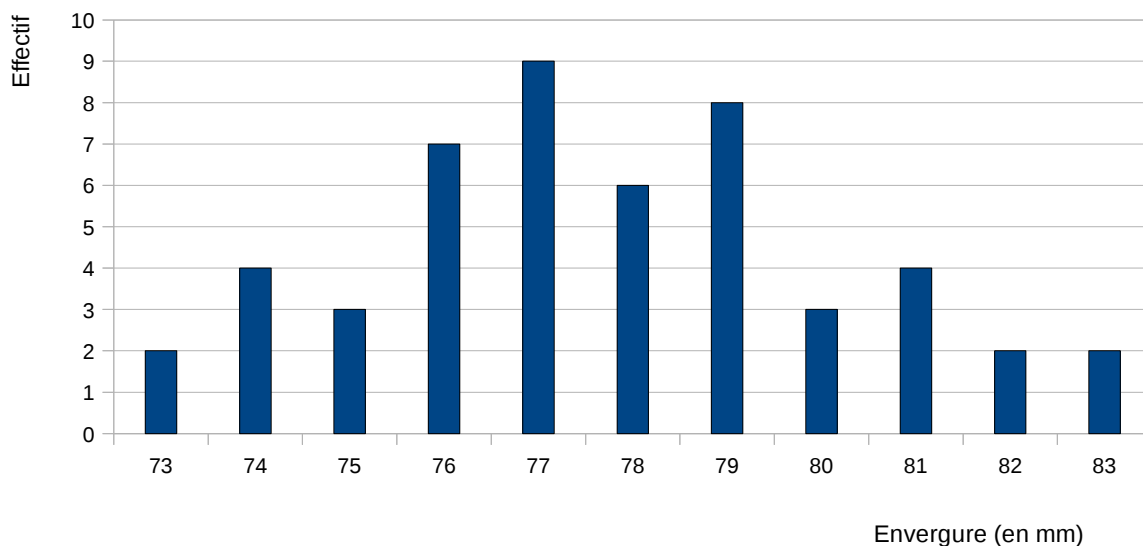
L'effectif de la série est  $n = n_1 + n_2 + \dots + n_r$

Valeurs prises	$x_1$	$x_2$	$x_3$	...	$x_r$
Effectifs	$n_1$	$n_2$	$n_3$	...	$n_r$

**Exemple :** On a dépouillé 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**.



## Fréquence

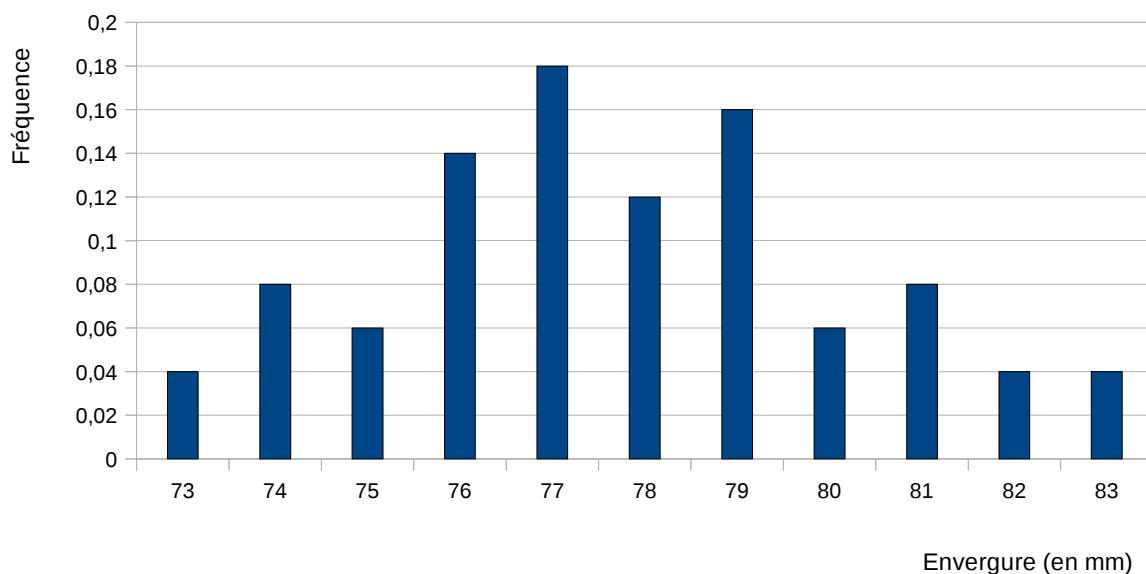
**Définition :** Une série peut être aussi définie par la **distribution de fréquences**, c'est-à-dire la liste des valeurs prises et la fréquence de chacune d'elles.

La somme des fréquences est 1

**Exemple :** 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

On obtient alors le diagramme en barre suivant :

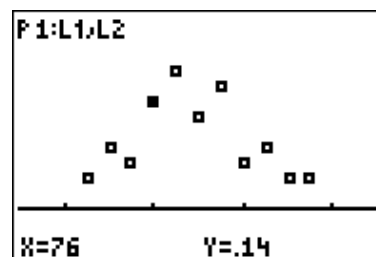


On peut également utiliser la calculatrice :

L1	L2	L3	2
73	.04	-----	
74	.08		
75	.06		
76	.14		
77	.18		
78	.12		
79	.16		
L2(1)= .04			

```

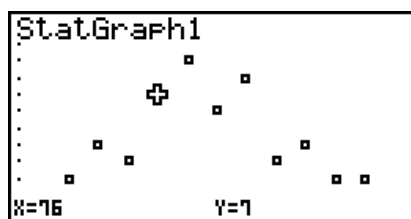
Graph1  Graph2  Graph3
--- NAff
Type: [ ] [ ] [ ]
ListeX: L1
ListeY: L2
Marque: [ ] +
  
```



SUB	List 1	List 2	List 3	List 4
2	74	4		
3	75	3		
4	76	7		
5	77	9		
GPH1 GPH2 GPH3 SEL SET				

```

StatGraph1
Graph Type : Scatter
XList      : List1
YList      : List2
Frequency   : 1
Mark Type   : [ ]
  
```



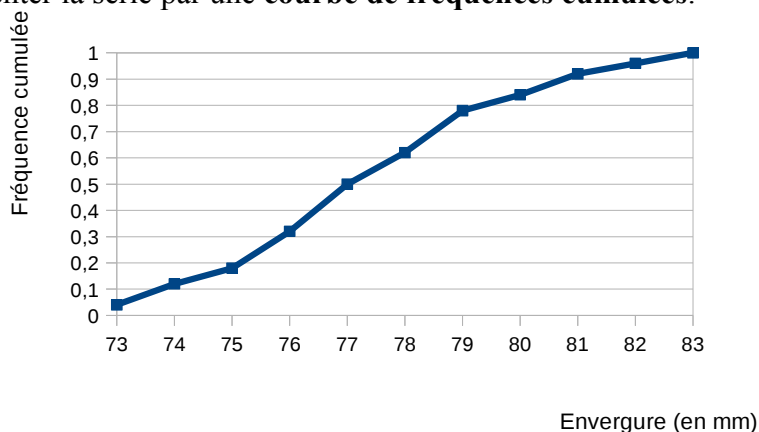
## 2) Fréquences cumulées

**Définition :** On peut compléter une distribution de fréquences par la ligne des **fréquences cumulées croissantes**, obtenues en ajoutant à chaque fréquence la somme des fréquences précédentes.

**Exemple :** 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**.



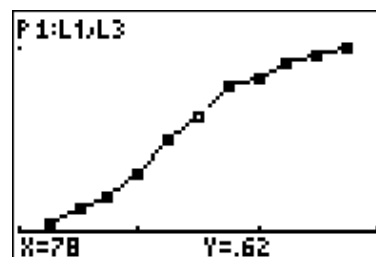
On peut également utiliser la calculatrice :

L1	L2	L3	3
73	.04	.04	
74	.08	.12	
75	.06	.18	
76	.14	.32	
77	.18	.5	
78	.12	.62	
79	.16	.78	

L3(1) = .04

```

Graph1 Graph2 Graph3
fff NAff
Type: [L1] [L2] [L3]
ListeX: L1
ListeY: L3
Marque: [ ] +
  
```



Sub	List 1	List 2	List 3	List 4
1	73	2		
2	74	4		
3	75	3		
4	76	7		

List 2 ÷ 50

Sub	List 1	List 2	List 3	List 4
1	73	2	0.04	
2	74	4	0.08	
3	75	3	0.06	
4	76	7	0.14	

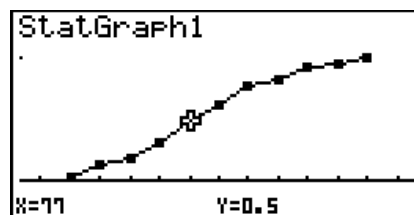
Cum1 List 3  
List L+M Dim Fill Seq

Sub	List 1	List 2	List 3	List 4
1	73	2	0.04	0.04
2	74	4	0.08	0.12
3	75	3	0.06	0.18
4	76	7	0.14	0.32

0.18  
LIST CLR CALC HYP PROB

```

StatGraph1
Graph Type : xyLine
XList : List1
YList : List4
Frequency : 1
Mark Type : [ ]
  
```



### 3) Regroupement par classes

Dans certains cas, on est amené à effectuer un regroupement des valeurs en classes.

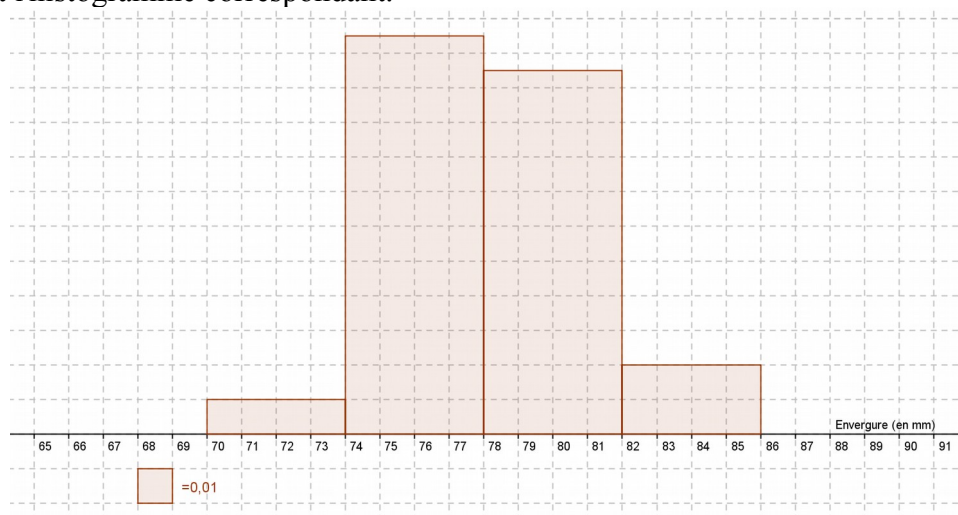
#### Définition :

- Un **histogramme** est un ensemble de rectangles adjacents ; chaque rectangle a une aire égale à la fréquence de la classe et une base égale à l'amplitude de la classe.  
L'aire totale de l'histogramme est 1.
- On peut également construire un **histogramme des effectifs** : chaque rectangle a une aire égale à l'effectif de la classe.  
L'aire totale est égale à l'effectif de la série.

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

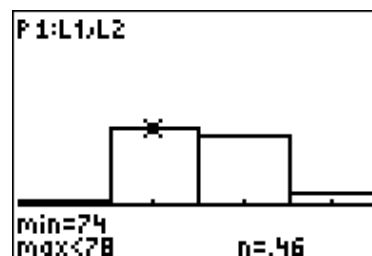
On construit l'histogramme correspondant.



Avec la calculatrice on obtient :

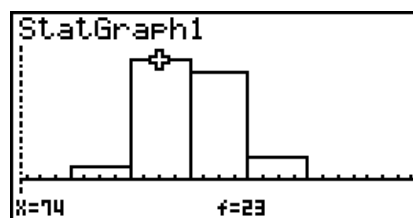
```
Graph1 Graph2 Graph3
MAff
Type: [ ] [ ] [ ]
ListeX:L1
Effectifs:L2
```

```
FENETRE
Xmin=70
Xmax=86
Xgrad=4
Ymin=-.2886624
Ymax=1.1232
Ygrad=1
Xres=1
```



```
StatGraph1
Graph Type :Hist
XList :List1
Frequency :List2
1 LIST
```

```
sur Réglage Histogram
intervalle
Start:70
Width:4
Dessin:[EXE]
```



Remarque : Si les classes ont toutes la même amplitude, les hauteurs des rectangles sont proportionnelles aux fréquences et l'histogramme est alors assimilable à un diagramme en bâtons.

### III. Paramètres d'une série

Pour résumer une **série quantitative**, on utilise des **caractéristiques de position** (moyenne, médiane, quartiles, etc...) et des **caractéristiques de dispersion** (étendue, écart interquartile, ...)

#### 1) Caractéristiques de position

##### Moyenne

**Propriété :** Pour déterminer la **moyenne** d'une série prenant les valeurs  $x_1, x_2, \dots, x_r$  avec les fréquences respectives  $f_1, f_2, \dots, f_r$ , on utilise la formule :

$$\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_r x_r$$

**Exemple :** Pour la série « envergure des papillons femelles », la moyenne calculée à partir de la distribution de fréquences est  $0,04 \times 73 + 0,08 \times 74 + 0,06 \times 75 + \dots + 0,04 \times 83 = 77,72$ .

L1	L2	L3	Z
73	.04	-----	
74	.08		
75	.06		
76	.14		
77	.18		
78	.12		
79	.16		
L2(1) = .04			

```
moyenne(L1,L2)
77.72
```

	List 1	List 2	List 3	List 4
SUB				
5	77	9	0.18	0.5
6	78	6	0.12	0.62
7	79	8	0.16	0.78
8	80	3	0.06	0.84
3				
GPH1 GPH2 GPH3 SEL SET				

```
Mean(List 1,List 2)
77.72
0
List L→M Dim Fill Seq ▸
```

##### Médiane et quartiles

###### Définitions :

- La **médiane** d'une série statistique est le nombre  $Me$  tel que :
  - 50% au moins des individus ont une valeur du caractère inférieure ou égale à  $Me$
  - 50% au moins des individus ont une valeur supérieure ou égale à  $Me$
- Le **premier quartile** est la plus petite valeur  $Q_1$  de la liste telle qu'au moins **un quart** des valeurs de la liste sont **inférieures ou égales** à  $Q_1$ .
- Le **troisième quartile** est la plus petite valeur  $Q_3$  de la liste telle qu'au moins les **trois quarts** des valeurs de la liste sont **inférieures ou égales** à  $Q_3$ .

### Remarques :

- Pour déterminer médiane et quartiles, il faut d'abord ordonner la série dans l'ordre croissant.

**Exemple :** Pour la série « envergure des papillons femelles »

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

Il y a 50 valeurs donc on prend la moyenne entre la 25<sup>ème</sup> et 26<sup>ème</sup> valeur : **Me = 77,5 mm**

$\frac{50}{4} = 12,5$  donc **Q<sub>1</sub>** correspond à la 13<sup>ème</sup> valeur.

**Q<sub>1</sub> = 76 mm**

$50 \times \frac{3}{4} = 37,5$  donc **Q<sub>3</sub>** correspond à la 38<sup>ème</sup> valeur.

**Q<sub>3</sub> = 79 mm**

- On peut également utiliser les **fréquences cumulées croissantes**.

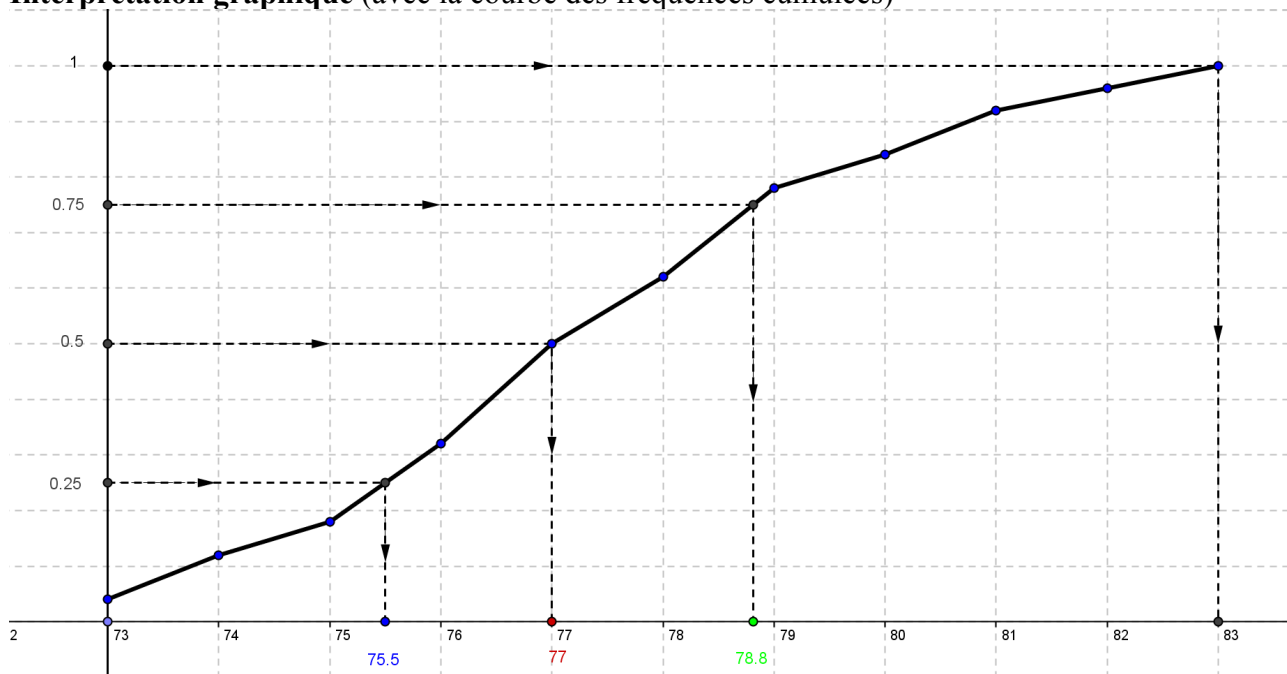
**Exemple :** Pour la série « envergure des papillons femelles »

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

La fréquence cumulée croissante 0,25 n'est pas atteinte ; la première valeur pour laquelle elle est dépassée est 76 : **Q<sub>1</sub> = 76 mm**.

La fréquence cumulée croissante 0,75 n'est pas atteinte ; la première valeur pour laquelle elle est dépassée est 79 : **Q<sub>3</sub> = 79 mm**.

### Interprétation graphique (avec la courbe des fréquences cumulées)



## 2) Caractéristiques de dispersion

### Définitions :

- L'**étendue** est la différence entre les valeurs extrêmes de la série.
- L'**écart interquartile** est la différence entre les 3<sup>ème</sup> et 1<sup>er</sup> quartiles ( $Q_3 - Q_1$ ).

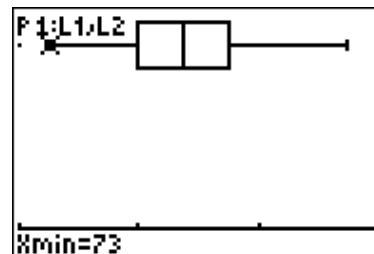
### Avec la calculatrice :

On peut obtenir un résumé de la série statistique.

On peut également utiliser les **boîtes à moustaches**.

```
Stats 1-Var
n=50
minX=73
Q1=76
Med=77.5
Q3=79
maxX=83
```

```
Graph1 Graph2 Graph3
Off NAff
Type: [L1] [L2] [L3]
ListeX:L1
Effectifs:L2
```



```
1Var XList :List1
1Var Freq :List2
2Var XList :List1
2Var YList :List2
2Var Freq :1
1 List
```

```
1 variable
x̄ =77.72
Σx =3886
Σx² =302332
σx =2.49831943
sx =2.52368373
n =50 ↓
```

```
StatGraph1
Graph Type :MedBox
XList :List1
Frequency :List2
Outliers :Off
```

