

# Statistiques

## 1) Vocabulaire :

**Population** : ensemble étudié ;

**Individu** : élément de la population ;

**Échantillon** : Partie de la population.

**Caractère (ou variable)** : ce que l'on étudie.

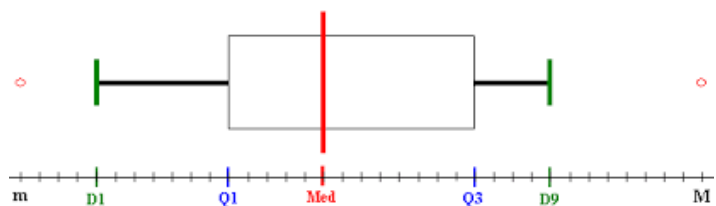
Un caractère peut être :

- **qualitatif** : les valeurs du caractère ne sont pas numériques (ex. : ville d'origine).
- **quantitatif** : les valeurs du caractère sont numériques (ex. taille).

Un caractère quantitatif peut être :

- **discret** : valeurs finies (ex. : nombre d'enfants) ;
- **continu** : valeurs infinies, on va créer des intervalles contigus , **appelés Classes.**

- Moyenne d'une série =
- l'écart type est un indicateur de dispersion des valeurs de la série,  
il est noté  $\sigma$  et on a  $\sigma = \sqrt{V}$
- avec  $V$  qui est la variance =
- La médiane  $Me$  d'une série est un nombre qui découpe la liste ordonnée de cette série en 2 listes de même effectifs.
- Les quartiles  $Q1$   $Me$  et  $Q3$  d'une série sont des nombres qui découpent une liste ordonnée en 4 listes de même effectifs .
- Les déciles  $D1, D2 \dots D9$  d'une série sont des nombres qui découpent une liste ordonnée en 10 listes de même effectifs.
- Diagramme en boîte :



|           |    |    |    |    |    |    |    |
|-----------|----|----|----|----|----|----|----|
| age       | 8  | 9  | 10 | 11 | 12 | 13 | 14 |
| effectifs | 12 | 15 | 20 | 18 | 12 | 10 | 8  |

|          |         |         |         |         |         |
|----------|---------|---------|---------|---------|---------|
| prix     | [10;20[ | [20;30[ | [30;40[ | [40;50[ | [50;60[ |
| effectif | 20      | 42      | 35      | 18      | 10      |
|          |         |         |         |         |         |

## 2) Variables qualitatives.

### 2.1) Tableau de contingence .

On présente généralement les données dans un tableau à double entrée, on parle de **tableau de contingence** qui permet d'effectuer des tris croisés.

#### Exemple :

Un sondage sur le type de musées le plus attractif a été réalisé sur un échantillon de 152 personnes. Les résultats sont les suivants :

|   | Enfants (E) | Adultes (A) | Total |
|---|-------------|-------------|-------|
| Beaux-arts (B)                            | 3           | 24          | 27    |
| Histoire -<br>Archéologie (H)             | 13          | 19          | 32    |
| Sciences (S)<br>(technologie, svt)        | 40          | 18          | 58    |
| Divers (D)<br>(automobiles,<br>jouet ...) | 25          | 10          | 35    |
| Total                                     | 81          | 71          | 152   |

Sur 71 adultes interrogés, 18 préfèrent les musées spécialisés en Sciences.

Sur ..... personnes interrogées, ..... préfèrent les musées spécialisés en Histoire

Sur ..... enfants interrogés, ..... préfèrent les musées spécialisés en Beaux arts

### 2.2) Fréquence conditionnelle.

Afin de pouvoir faire des comparaisons, on utilise des fréquences, généralement exprimées en pourcentage.

**On note  $f_A(S)$  la fréquence de S parmi A, on dit " S sachant A "**

$$f_A(S) = \frac{\text{Effectif de "A et S"}}{\text{Effectif de A}} = \frac{18}{71} \approx 0,254 \quad \text{soit } 25,4 \%$$

Parmi les adultes, 25,4 % préfèrent les musées de sciences.

Chercher les 2 fréquences conditionnelles correspondantes aux 2 exemples restants.

### 3)- Variables quantitatives

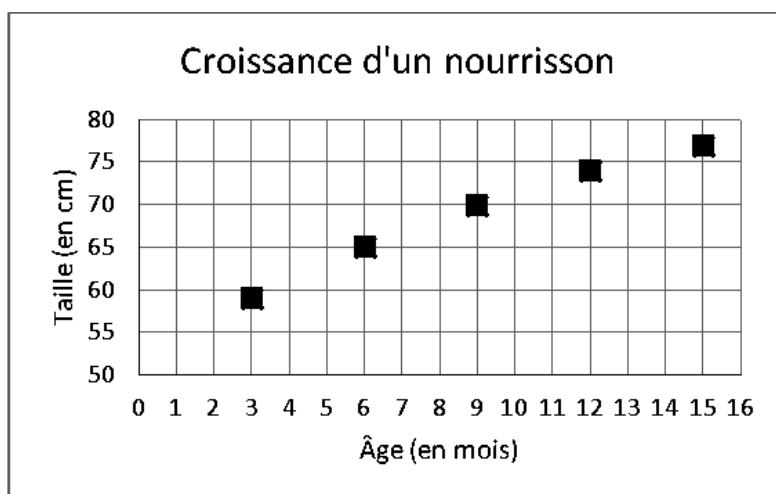
#### 3.1) Nuage de points :

On note généralement les deux variables étudiées  $x_i$  et  $y_i$ , ce qui permet une représentation graphique **par nuage de points** dans un repère ( O, x ; y) habituel.

#### Exemple :

On étudie la taille d'un nourrisson et son âge, on obtient une série statistique à 2 variables :

|                   |    |    |    |    |    |
|-------------------|----|----|----|----|----|
| Âge (en mois) $x$ | 3  | 6  | 9  | 12 | 15 |
| Taille (cm) $y$   | 59 | 65 | 70 | 74 | 77 |



#### 3.2) Point moyen.

Le point moyen est noté **G**, ses coordonnées sont données par la moyenne des abscisses et la moyenne des ordonnées.

$$\bar{x} = \frac{3+6+9+12+15}{5} = 9 \quad \text{et} \quad \bar{y} = \frac{59+65+70+74+77}{5} = 69$$

Utilisation de la calculatrice :

Menu Stat : On saisit les valeurs de x en list1 et celles de y en list2.

|     | List 1 | List 2 | List 3 | List 4 |
|-----|--------|--------|--------|--------|
| SUB |        |        |        |        |
| 1   | 3      | 59     |        |        |
| 2   | 6      | 65     |        |        |
| 3   | 9      | 70     |        |        |
| 4   | 12     | 74     |        |        |

```
2-Variable
Σx² = 495
xσn = 4.24264068
xσn-1 = 4.74341649
n = 5
ȳ = 69
Σy = 345
```

CALC puis 2var permet d'obtenir (en autres) les valeurs de  $\bar{x}$  et  $\bar{y}$

### 3.3) Ajustement à l'aide d'une fonction.

Effectuer un ajustement d'un nuage de points, c'est déterminer une fonction dont la représentation graphique passe « au plus près » de tous les points du nuage.

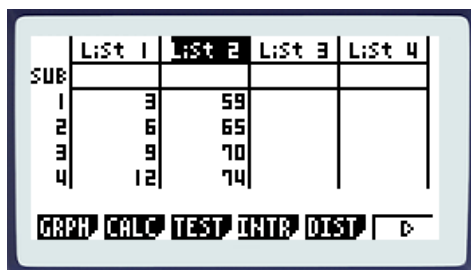
Cas particulier : Ajustement affine à l'aide d'une droite.

Lorsque les points semblent alignés, on cherche une droite d'équation  $y = ax + b$ .

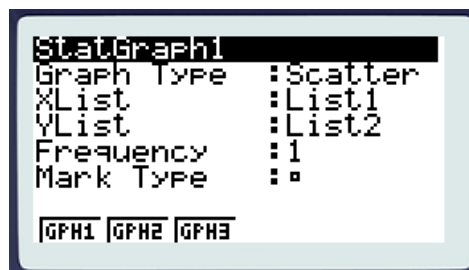
Cette droite passe par le point moyen.

Remarque : La calculatrice propose un ajustement affine :

Après avoir saisi les données en list1 et list2 dans le menu Stat.



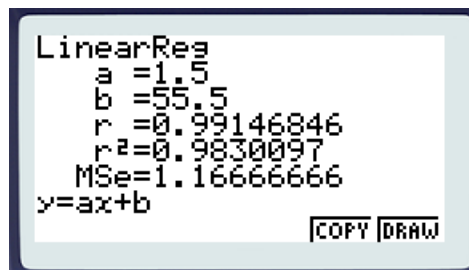
| SUB | List 1 | List 2 | List 3 | List 4 |
|-----|--------|--------|--------|--------|
| 1   | 3      | 59     |        |        |
| 2   | 6      | 65     |        |        |
| 3   | 9      | 70     |        |        |
| 4   | 12     | 74     |        |        |



```
StatGraph1
Graph Type : Scatter
XList      : List1
YList      : List2
Frequency  : 1
Mark Type  : □
|GPH1 |GPH2 |GPH3
```

GRPH puis SET pour sélectionner le type de graphique : **Scatter**

Sortir avec " Exit " et choisir GPH1.

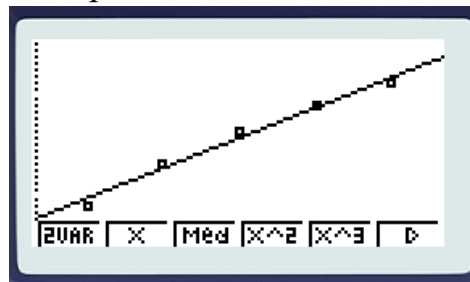


```
LinearReg
a = 1.5
b = 55.5
r = 0.99146846
r^2 = 0.9830097
MSE = 1.166666666
y = ax + b
|COPY |DRAW
```

Choisir ensuite CALC puis X : On obtient les valeurs de  $a$  et  $b$ .

ici  $y = 1,5x + 55,5$

Draw permet de tracer la droite



### **3.4) Estimations.**

On utilise la relation donnée par l'ajustement pour faire des estimations :

- Si on estime une valeur « entre deux points du nuage », on dit qu'on fait une **interpolation**,
- Lorsque cette estimation est faite « à l'extérieur du nuage », on dit qu'on fait une **extrapolation**.

Dans l'exemple , on peut ainsi estimer la taille d'un nourrisson de 18 mois  
en calculant  $y = 1,5 \times 18 + 55,5 = 82,5$  cm.