# Statistiques

## I Rappels

### 1 - Série statistique à une variable

Définition :

Une **série statistique à une variable** $(x_i)$ est constituée d'une liste de valeurs $x_1$ ; $x_2$ ; … ; $x_p$ avec les effectifs correspondants $n_1$ ; $n_2$ ; … ; $n_p$

Exemple : Les 31 élèves d'une classe de terminale ont obtenu les notes suivantes à une évaluation de mathématiques.

Valeurs $x_i$

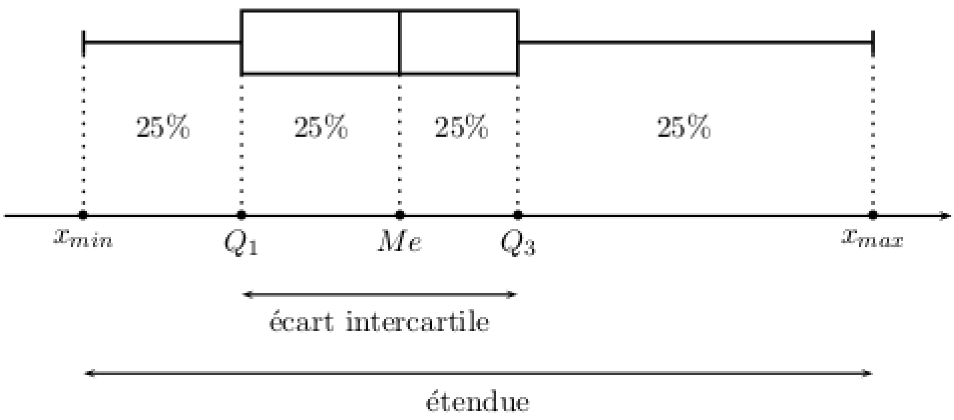
| Notes | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|
| Total | 1 | 5 | 4 | 12 | 5 | 3 | 0 | 1 |

Effectifs $n_i$

Définition :

Les valeurs de la série étant rangées dans l'ordre croissant, les **quartiles $Q_1$** et **$Q_3$** et la **médiane $Me$** partagent la série ordonnée en quatre groupes de même effectif.

Schéma : On peut représenter cette répartition par un diagramme en boîte.




Paramètres d'une série statistique :

- **La moyenne** : $\overline{x} = \frac{x_1+x_2+\cdots+x_n}{N}$
- **La médiane** $Me$ partage la série ordonnée en deux groupes de même effectif.
  Si l'effectif est impair, la médiane est la valeur située au milieu de la série.
  Si l'effectif est pair, la médiane est la moitié de la somme des deux valeurs au milieu.
- **L'étendue** : c'est la différence des valeurs extrêmes : $e = x_{max} - x_{min}$
- **L'écart-type** mesure la dispersion par rapport à la moyenne. On le note $\sigma$.
- **Les quartiles $Q_1$ et $Q_3$**
- **L'écart interquartile $[Q_1 \; ; \; Q_3]$**

Application : Exercice 1

## 2- Fonction affine et équation réduite de droite

Définition :

Une **fonction affine** $f$ est une fonction définie pour tout réel $x$ par :

$$f(x) = ax + b$$

Le nombre $a$ est le **coefficient directeur** (ou pente).

Le nombre $b$ est **l'ordonnée à l'origine**.

Propriétés :

Si $a = 0$ alors la fonction est **constante** et s'écrit sous la forme $f(x) = b$

Si $b = 0$ alors la fonction est **linéaire** et s'écrit sous la forme $f(x) = ax$

Propriété :

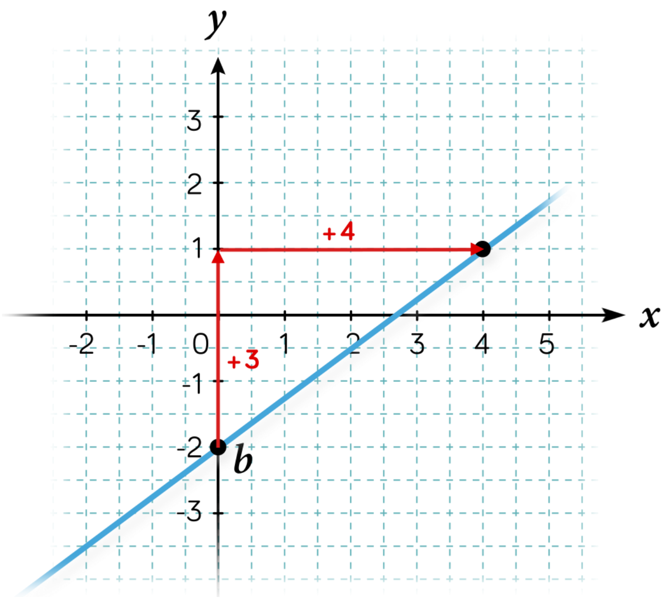
La représentation graphique d'une fonction affine est une droite.

Méthode : Déterminer graphiquement l'équation de la droite représentative d'une fonction affine

Exemple : Soit $f$ une fonction affine dont la représentation graphique est donnée ci-contre.

- Pour déterminer $b$, il suffit de regarder le point d'intersection entre la droite et l'axe des ordonnées. Ici, $b = -2$
- Pour déterminer $a$, il faut observer le nombre de carreaux dont on se déplace horizontalement et verticalement pour partir de $b$ et arriver à un point avec des coordonnées « entières ».

$$a = \frac{déplacement\ vertical}{déplacement\ horizontal} = \frac{3}{4}$$




Définition :

Soient $A(x_A\ ;\ y_A)$ et $B(x_B\ ;\ y_B)$ deux points distincts de la droite représentative d'une fonction affine $f$ telle que $f(x) = ax + b$.

Le coefficient directeur $a$ est égal à : $a = \frac{f(x_B)-f(x_A)}{x_B-x_A} = \frac{y_B-y_A}{x_B-x_A}$

Méthode : Déterminer par le calcul l'équation de la droite représentative d'une fonction affine

Soient $A(1\,;2)$ et $B(6\,;8)$ deux points d'une fonction affine $f$ et $d$ sa droite représentative.

On veut déterminer par le calcul l'expression de la fonction $f$.

- On commence par la valeur du coefficient directeur de la fonction $f$ :

$$a = \frac{y_B - y_A}{x_B - x_A}$$
$$a = \frac{8-2}{6-1}$$
$$a = \frac{6}{5}$$

Donc : $f(x) = \frac{6}{5}x + b$

- On peut ensuite déterminer $b$ en remplaçant $x$ et $f(x)$ dans l'expression de la fonction par les coordonnées d'un point appartenant à la droite $d$.

$$f(x_A) = \frac{6}{5}x_A + b$$
$$2 = \frac{6}{5} \times 1 + b$$
$$2 = \frac{6}{5} + b$$
$$2 - \frac{6}{5} = b$$
$$\frac{4}{5} = b$$

Donc : $f(x) = \frac{6}{5}x + \frac{4}{5}$

Application : Exercice 2

# II Série statistique à deux variables

## 1 - Définitions

Définition :

Une **série statistique à deux variables** $(x_i\,;\,y_i)$ est constituée d'une liste de $n$ couples de valeurs $(x_1\,;\,y_1)\,;\,(x_2\,;\,y_2)\,;\ldots;(x_n\,;\,y_n)$.

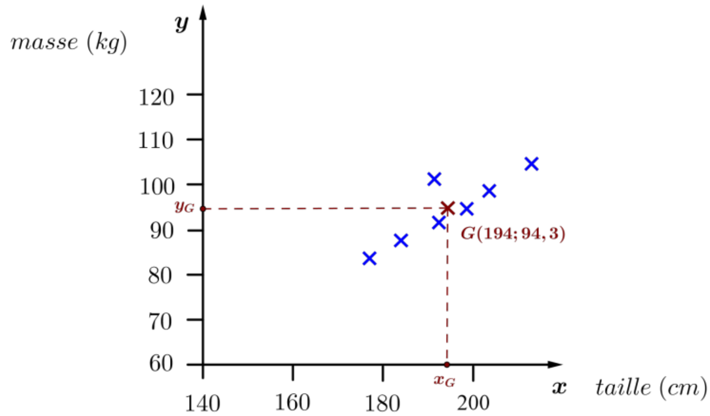
Le **nuage de points** de cette série est l'ensemble des points du plan de coordonnées $(x_1\,;\,y_1)\,;\,(x_2\,;\,y_2)\,;\ldots;(x_n\,;\,y_n)$.

Remarque : Le nuage de points permet de visualiser un lien possible entre deux variables ou de repérer des incohérences dans une série statistique à deux variables.

Exemple : Le tableau suivant regroupe la taille et la masse de sept joueurs

| Taille (cm) | 183 | 192 | 177 | 213 | 191 | 199 | 203 |
|---|---|---|---|---|---|---|---|
| Masse (kg) | 88 | 91 | 83 | 105 | 101 | 95 | 97 |

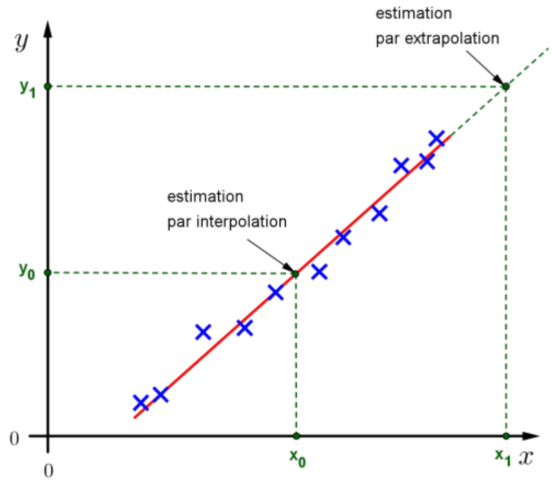
La représentation graphique est la suivante :




Les **coordonnées du point moyen *G*** sont $(\bar{x}\ ;\ \bar{y})$, moyenne des $x_i$, et moyenne des $y_i$.

Le nuage de points permet de constater qu'un des joueurs semble être en surcharge pondérale par rapport aux autres.

<u>Application</u> : Exercice 3

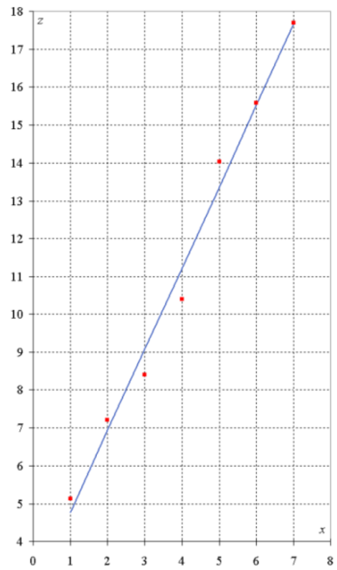



<u>Remarque</u> : Pour estimer une valeur inconnue dans une série statistique, on peut utiliser deux procédés :

- La méthode par interpolation quand le calcul est réalisé dans l'intervalle des valeurs prises par la série statistique.
- La méthode par extrapolation quand le calcul est réalisé en dehors des valeurs prises par la série statistique.

## 2- Droite d'ajustement

### a- Méthode « au jugé »

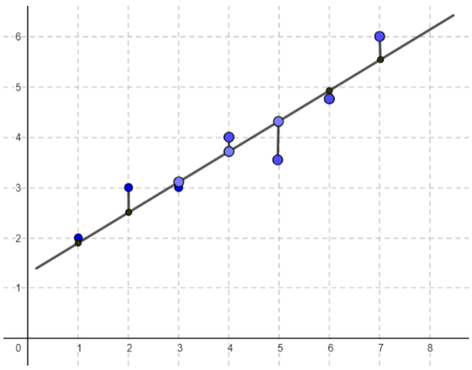


La méthode « au jugé » consiste à tracer une droite passant au plus près des différents points du nuage.

Après avoir tracé cette droite, on peut en déterminer son équation graphiquement ou par le calcul en déterminant les coordonnées de deux points appartenant à la droite (Méthodes précédentes).

<u>Application</u> : Exercice 4

### b- Méthode des moindres carrés



L'équation de la droite d'ajustement peut être aussi déterminée à l'aide d'une calculatrice ou d'un tableur, par la <u>**méthode des moindres carrés**</u>.

[Vidéo explicative et tutos Casio/TI]

[Tuto Numworks]

### c- Méthode des points moyens (ou droite de Mayer)

Cette méthode consiste à déterminer deux points moyens du nuage de points et de déterminer par le calcul l'équation de la droite passant par ces deux points moyens.

Exemple : On considère la série statistique suivante

| Superficie $X$ | 2 | 2 | 3 | 4 | 5 | 6 | 7 | 7,6 |
|---|---|---|---|---|---|---|---|---|
| Nombre d'exploitations $Y$ | 14 | 26 | 31 | 29 | 44 | 40 | 54 | 50 |

Les valeurs du caractère $X$ sont 8 au total. On va donc partager la série en deux séries de 4 valeurs :

Série 1

| $X$ | 2 | 2 | 3 | 4 |
|---|---|---|---|---|
| $Y$ | 14 | 26 | 31 | 29 |

Point moyen $G_1\ (\overline{X_1}\ ;\ \overline{Y_1})$ :

$$\overline{X_1} = \frac{2+2+3+4}{4} = 2{,}75$$

$$\overline{Y_1} = \frac{14+26+31+29}{4} = 25$$

Série 2

| $X$ | 5 | 6 | 7 | 7,6 |
|---|---|---|---|---|
| $Y$ | 44 | 40 | 54 | 50 |

Point moyen $G_2\ (\overline{X_2}\ ;\ \overline{Y_2})$ :

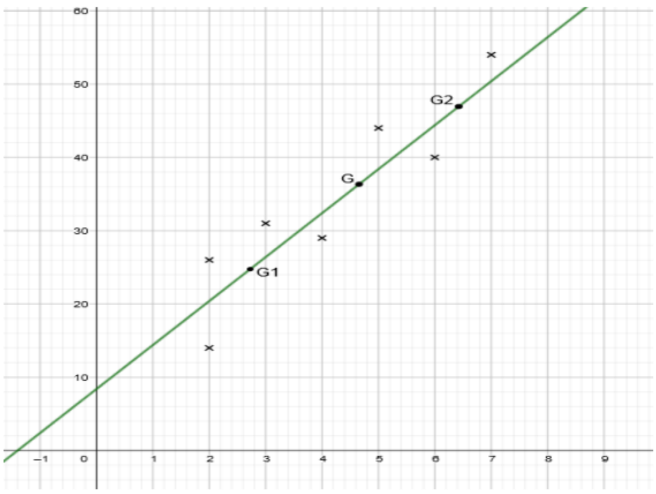
$$\overline{X_2} = \frac{5+6+7+7{,}6}{4} = 6{,}4$$

$$\overline{Y_2} = \frac{44+40+54+50}{4} = 47$$

On peut ensuite déterminer l'équation de la droite $(G_1G_2)$ grâce à la méthode vue précédemment pour déterminer l'équation réduite d'une droite à l'aide des coordonnées des deux points $G_1$ et $G_2$ :

$$a = \frac{47-25}{6{,}4-2{,}75} \approx 6 \qquad b = 25 - 6 \times 2{,}75 = 8{,}4$$

Donc : l'équation de la droite d'ajustement $(G_1G_2)$ est $y = 6x + 8{,}4$




Application : Exercice 5

## 3- Ajustement par changement de variable

Il arrive que le nuage de points ne soit pas modélisable par une droite. Il faut alors changer de variable pour réaliser un ajustement linéaire.

Exemple : Un centre d'appel comptait 66 employés en 2001. Le tableau ci-dessous donne l'évolution du nombre d'employés en fonction du range de l'année.

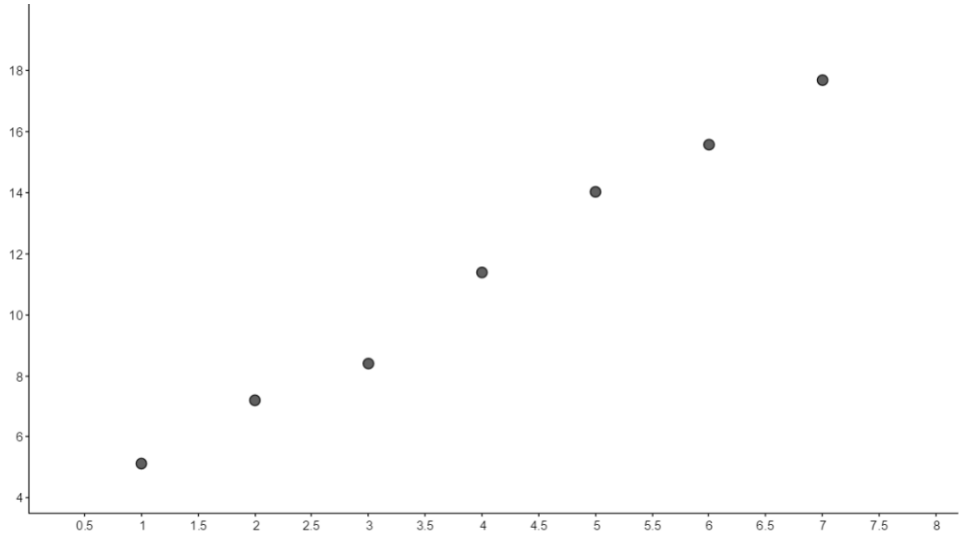
| Année | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| Rang de l'année $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Nombre d'employés $y_i$ | 66 | 104 | 130 | 207 | 290 | 345 | 428 |

On cherche à étudier l'évolution du nombre $y$ d'employés en fonction du rang $x$ de l'année. Une étude graphique montre qu'un ajustement affine ne convient pas.

On pose alors $z = \sqrt{y} - 3$

| Année | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| Rang de l'année $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Nombre d'employés $y_i$ | 66 | 104 | 130 | 207 | 290 | 345 | 428 |
| $z_i$ | 5,12 | 7,2 | 8,4 | 11,39 | 14,03 | 15,57 | 17,69 |

On obtient le graphique suivant après le changement de variable. Les points semblent presque alignés, ce qui rend réalisable un ajustement affine (linéaire) par l'une des méthodes vues précédemment.




<u>Application</u> : Exercice 6