

Statistique

1 Série statistique à une variable (rappels)

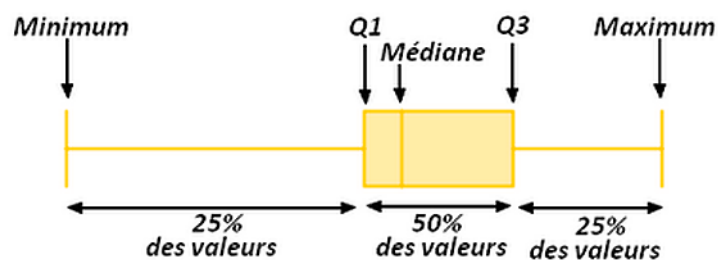
Définition 1 – Série statistique à une variable

Une **série statistique à une variable** (x_i) est constituée d'une liste de n valeurs $x_1 ; x_2 ; \dots ; x_n$.

Définition 2 – Quartiles et Médiane

Les valeurs de la série étant rangées dans l'ordre croissant, les **quartiles Q1 et Q3** et la **médiane** partagent la série ordonnée en quatre groupes de même effectif.

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



Définition 3 – Paramètres d'une série statistique

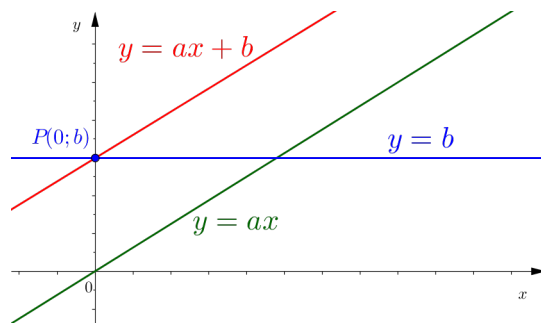
Une série statistique peut se résumer par quelques paramètres :

- la **moyenne** \bar{X}
- l'**écart type** σ
- l'**intervalle** $[\bar{X} - \sigma; \bar{X} + \sigma]$.
- la **médiane** Me
- les **quartiles Q1 et Q3**
- l'**intervalle interquartile** [Q1 ; Q3]

2 Fonction affine et droite

Définition 4 – Fonction affine

Une **fonction affine** est une fonction définie sur \mathbb{R} par $f(x) = ax + b$, où a et b sont des nombres réels.



Remarque : Une fonction affine est représentée par une droite \mathcal{D} de **coefficient directeur** a . Cette droite passe par le point $P(0; b)$, donc lorsque $b = 0$ la droite passe par l'origine (la fonction est linéaire).

Lorsque $a = 0$, alors $f(x) = b$ et la fonction est constante sur \mathbb{R} : la droite est horizontale.

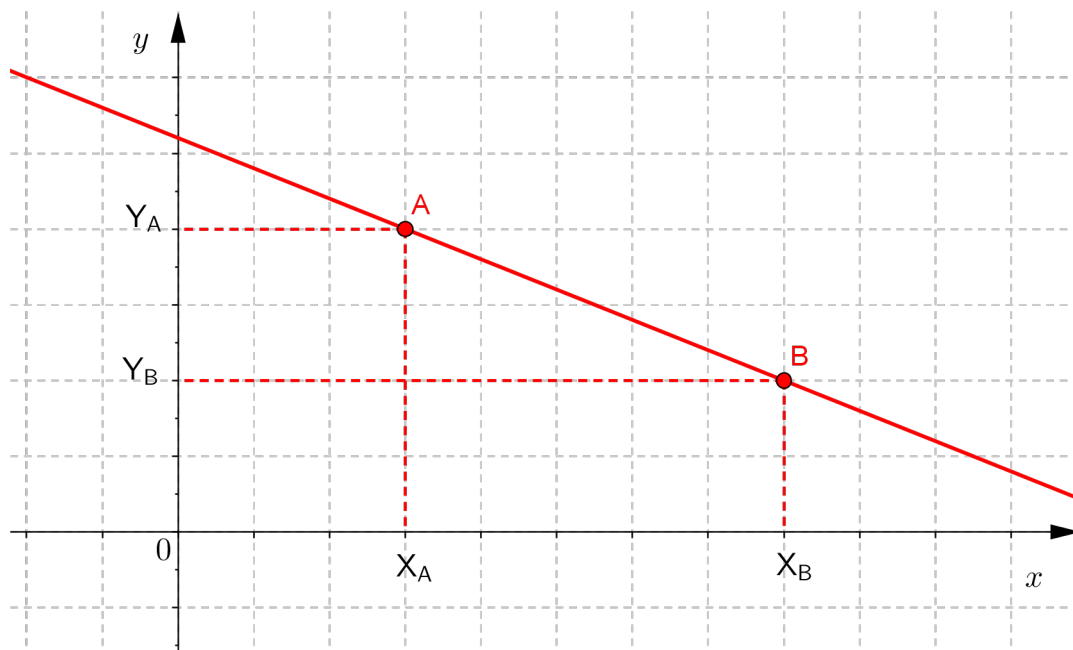
3 Équation réduite d'une droite

Soit une droite \mathcal{D} passant par deux points $A(X_A; Y_A)$ et $B(X_B; Y_B)$, avec $X_A \neq X_B$.

- La droite \mathcal{D} représente une fonction affine de coefficient directeur :

$$a = \frac{Y_B - Y_A}{X_B - X_A} = \frac{\text{différence des ordonnées}}{\text{différence des abscisses}}$$

- Son **équation réduite** est de la forme $y = a(x - X_A) + Y_A$



Méthode : La connaissance des coordonnées de deux points appartenant à une droite est suffisante pour déterminer l'équation réduite de cette droite.

4 Séries statistiques à deux variables

Définition 5 – Séries statistiques à deux variables

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

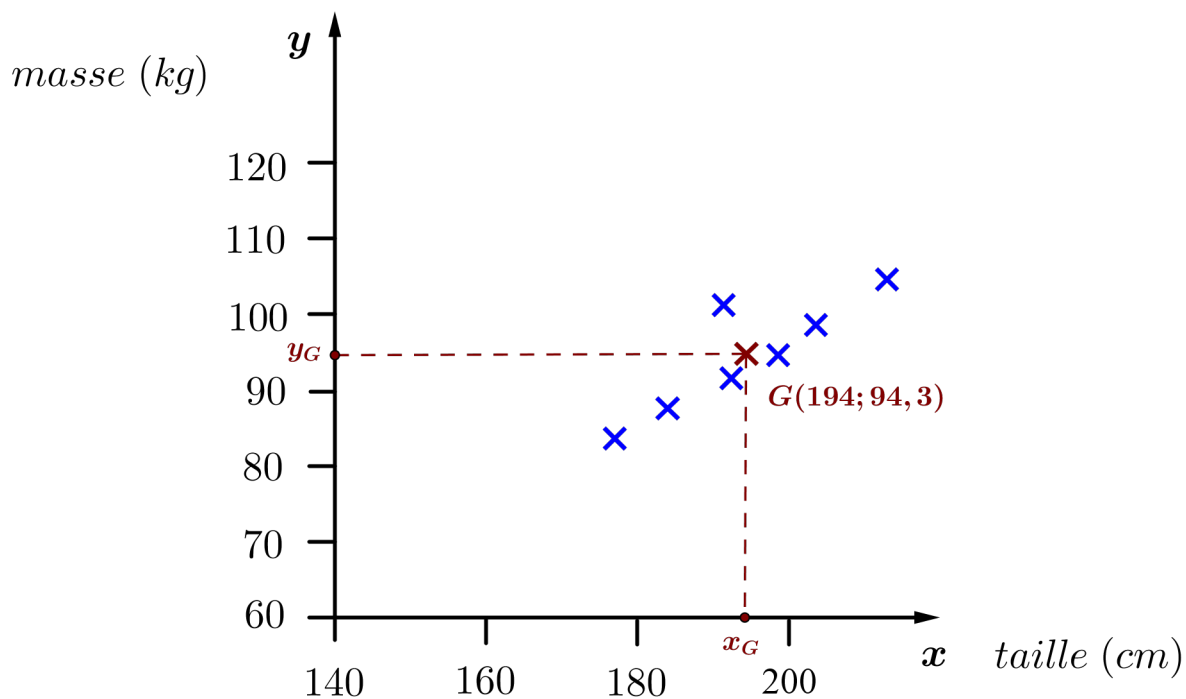
Définition 6 – Nuage de points

Soit une série statistique à deux variables $(x_1; y_1)$, $(x_2; y_2)$, ..., $(x_n; y_n)$. Le **nuage de points** de cette série est l'ensemble des n points du plan de coordonnées $(x_1; y_1)$, $(x_2; y_2)$, ..., $(x_n; y_n)$.

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

Exercice 1 : Le tableau suivant donne la taille et la masse de sept joueurs de basket.

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



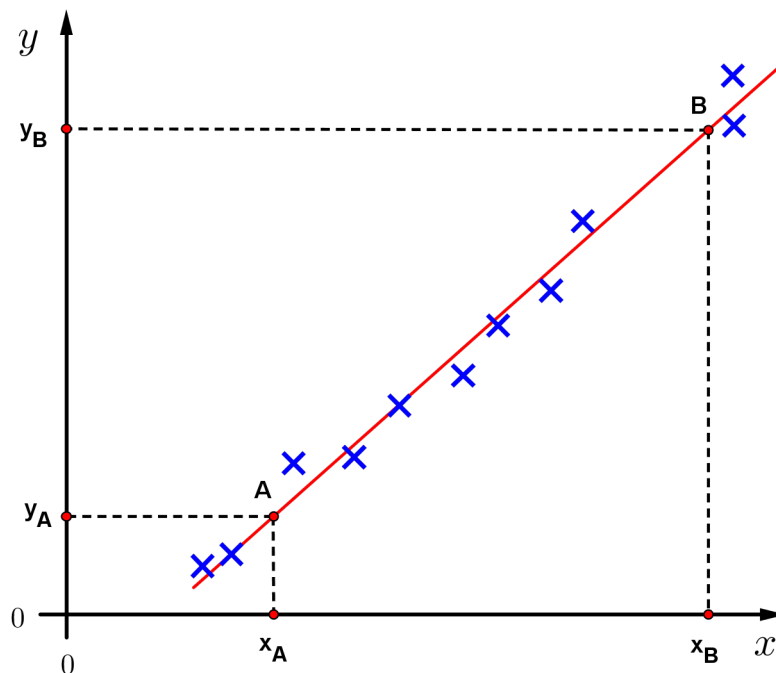
Les **coordonnées du point moyen G** sont $(\bar{x}; \bar{y})$, moyenne des x_i et moyenne des y_i .

Le nuage de point permet de constater qu'un des joueurs semble en surcharge pondérale par rapport aux autres joueurs de l'équipe.

5 Ajustement affine

Définition 7

Lorsque les points du nuage sont **presque alignés**, on peut approcher le nuage de point par une droite, représentation graphique d'une **fonction affine** exprimant y en fonction de x . Cette droite est appelée **droite d'ajustement**



5.1 Équation de la droite d'ajustement

L'équation de la droite d'ajustement peut être déterminée par une **méthode graphique** ou avec la **calculatrice**.

Méthode (détermination graphique) : On peut choisir deux points $A(x_A; y_A)$ et $B(x_B; y_B)$. La droite passant par A et B a pour coefficient directeur le nombre :

$$a = \frac{y_B - y_A}{x_B - x_A}$$

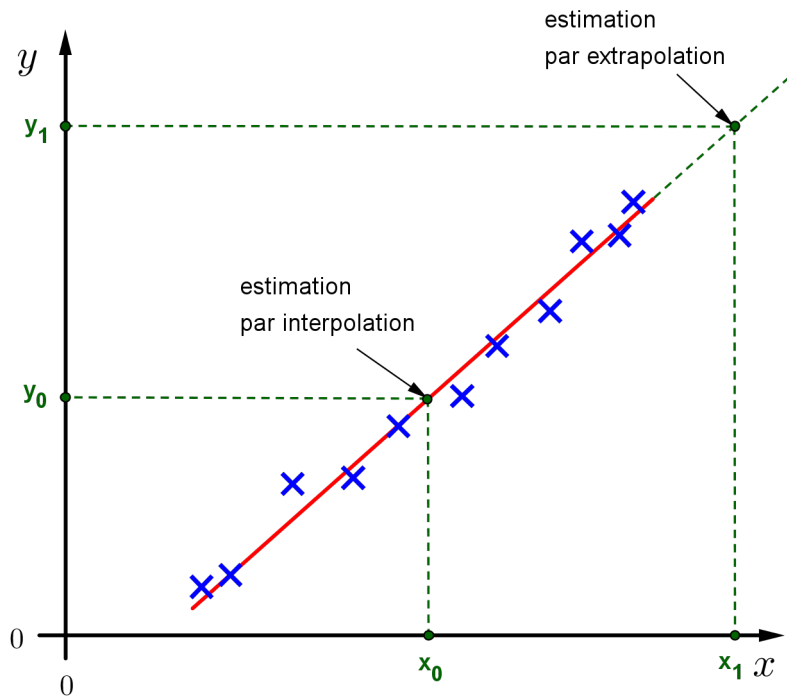
Son équation réduite est de la forme $y = a(x - x_A) + y_A$.

Méthode (méthode des moindres carrés avec la calculatrice) : à l'aide de la calculatrice ou du tableur, on obtient les coefficients a et b de l'équation $y = ax + b$.

5.2 Estimation par interpolation ou par extrapolation

Méthode (interpolation) : Pour une valeur donnée x_0 appartenant à l'intervalle des valeurs prises par x on peut estimer la valeur correspondante y_0 de la variable y .

Méthode (extrapolation) : Pour une valeur donnée x_1 n'appartenant pas à l'intervalle des valeurs prises par x on peut estimer la valeur correspondante y_1 de la variable y (on suppose que la relation affine entre x et y se prolonge au delà des valeurs étudiées).



5.3 Séries chronologiques

Définition 8 – Série chronologique

Une **série chronologique** est une série statistique dont la variable x est le temps.

Définition 9 – Préviation

Une **préviation** est une estimation par extrapolation d'une valeur x ou y d'une série chronologique.

