# 14 Statistique descriptive

Table des matières



## 14.1 Statistique univariée

Dans ce chapitre, pour illustrer les notions du cours nous nous appuierons, entres autres, sur la série suivante (longueur en cm de feuilles d'un plant de maïs) :

| Longueur $x_i$ (cm) | Effectif $n_i$ |
|---|---|
| 8 | 1 |
| 9 | 2 |
| 10 | 3 |
| 11 | 5 |
| 12 | 4 |
| 13 | 3 |
| 14 | 2 |
| **Total** | **20** |

### 14.1.1 Définitions

Une étude statistique univariée consiste à traiter des données mesurées, qu'elles soient quantitatives (âge, taille, poids, précipitations, températures, longueurs, sommes d'argent...) ou qualitatives (sexe, couleur, opinion...). Ces données sont appelées des caractères, ou des variables.

**Définition 1**

- L'ensemble d'objets ou de personnes sur lequel porte une étude statistique est appelé population.
- Un élément de cette population est appelé un individu.
- L'effectif d'une population est le nombre d'individus total de cette population.
- La fréquence d'un caractère est le nombre d'individus possédant ce caractère divisé par l'effectif total de la population.
- Une variable est dite discrète si elle ne prend que des valeurs isolées (par exemple des nombres entiers, comme pour les âges).
- Une variable est dite continue si elle peut prendre toutes les valeurs d'un intervalle (comme pour les précipitations, la température, ou plus généralement toute mesure d'une grandeur physique).

**Exemple 1.** • Le caractère étudié est la longueur des feuilles de maïs. Il s'agit d'une variable quantitative continue (même si le tableau ne présente que des valeurs arrondies à l'entier près).

- L'effectif de la longueur 10 est 3. Sa fréquence est $\frac{3}{20} = 15\%$.
- L'effectif total de cette série est 20.

### 14.1.2 Description d'une série statistique discrète

Considérons un effectif de taille $n$ et une série statistique $(x_1, \dots, x_n)$ portant sur un caractère discret $x$, où pour tout $i \in [\![1, n]\!]$, $x_i$ désigne la valeur du caractère mesurée pour l'individu $i$.

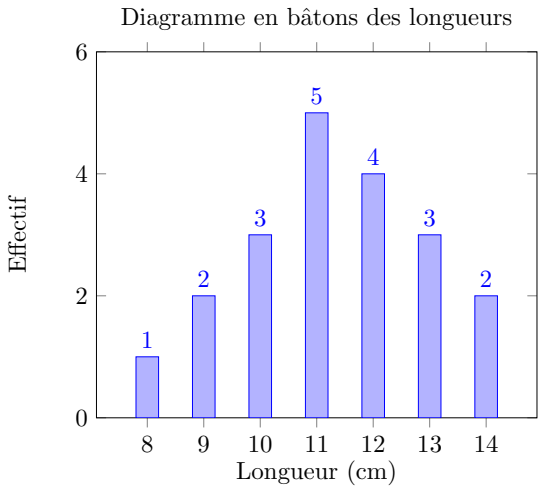
On peut présenter les données sous forme d'un tableau en regroupant les différentes valeurs du caratère prises par la population : si $\{x_1, \dots, x_n\} = \{y_1, \dots, y_p\}$ avec $y_1 < y_2 < \dots < y_p$, alors on peut écrire

| Caractère | Effectif |
|---|---|
| $y_1$ | $n_1$ |
| $y_2$ | $n_2$ |
| $\dots$ | $\dots$ |
| $y_p$ | $n_p$ |

où pour tout $i \in [\![1, p]\!]$, $n_i$ est le nombre d'individus pour lesquels le caractère prend la valeur $y_i$ et $\sum_{k=1}^{p} n_k = n$.

On représente généralement ces résultats sous la forme d'un diagramme à bâtons.

**Exemple 2.** Diagramme à bâtons de la longueur des feuilles de maïs :




**Définition 2: Fréquence**

Avec les notations précédentes, la fréquence du caractère $y_i$ est $f_i = \dfrac{n_i}{n}$.

**Remarque 1.** On a alors $\displaystyle\sum_{i=1}^{p} f_i = 1$.

**Exemple 3.** Tableau des fréquences (en %) des longueurs de feuilles de maïs :

| Longueur $x_i$ (cm) | Effectif $n_i$ | Fréquence (%) |
|---|---|---|
| 8 | 1 | 5 |
| 9 | 2 | 10 |
| 10 | 3 | 15 |
| 11 | 5 | 25 |
| 12 | 4 | 20 |
| 13 | 3 | 15 |
| 14 | 2 | 10 |
| **Total** | **20** | **100** |

**Exemple 4.** Dans une classe de 15 élèves, on relève les notes suivantes au dernier devoir de mathématiques :

| Note | Nombre d'élèves |
|---|---|
| 10 | 4 |
| 12 | 6 |
| 14 | 4 |
| 18 | 1 |

On a alors les fréquences suivantes :

| Note | Fréquence |
|---|---|
| 10 | $\frac{4}{15}$ |
| 12 | $\frac{2}{5}$ |
| 14 | $\frac{4}{15}$ |
| 18 | $\frac{1}{15}$ |

**Définition 3: Effectifs et fréquences cumulés croissants**

• Avec les notations précédentes, on définit pour tout $i \in [\![1, p]\!]$ les effectifs cumulés croissants

$$N_i = \sum_{k=1}^{i} n_i.$$

Autrement dit, $N_i$ représente le nombre d'invidus pour lesquels la valeur du caractère est inférieure (ou égale) à $y_i$.

• On définit de même les fréquences cumulés croissants

$$F_i = \sum_{k=1}^{i} f_i.$$

**Exemple 5.** Avec l'exemple précédent (notes au dernier devoir de mathématiques) , on obtient

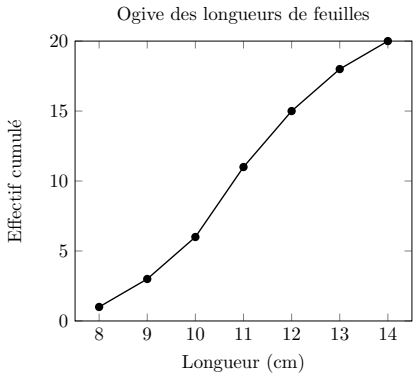
| Note | Effectifs cumulés croissants |
|---|---|
| 10 | 4 |
| 12 | 10 |
| 14 | 14 |
| 18 | 15 |

| Note | Fréquences cumulées croissantes |
|---|---|
| 10 | $\frac{4}{15}$ |
| 12 | $\frac{2}{3}$ |
| 14 | $\frac{14}{15}$ |
| 18 | 1 |

**Exemple 6.** Voici le tableau des effectifs et fréquences cumulés pour la longueur des feuilles de maïs :

| Longueur $x_i$ (cm) | Effectif cumulé $N_i$ | Fréquence cumulée (%) |
|---|---|---|
| 8 | 1 | 5 |
| 9 | 3 | 15 |
| 10 | 6 | 30 |
| 11 | 11 | 55 |
| 12 | 15 | 75 |
| 13 | 18 | 90 |
| 14 | 20 | 100 |

On peut également représenter la courbe des effectifs (ou fréquences) cumulé(e)s :

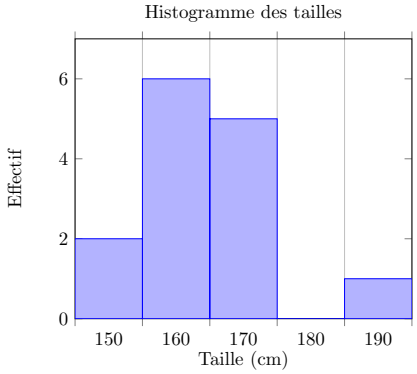



### 14.1.3 Description d'une série statistique continue

Prenons l'exemple de la taille d'élèves dans une classe de 14 élèves.

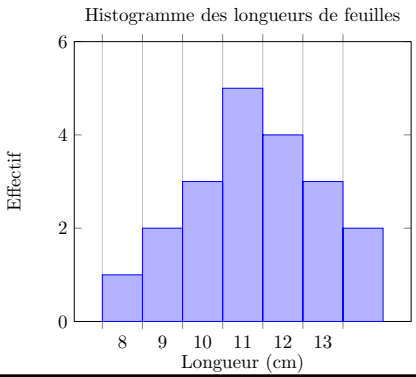
| Taille (cm) | Effectif |
|---|---|
| $[150, 160[$ | 2 |
| $[160, 170[$ | 6 |
| $[170, 180[$ | 5 |
| $[180, 190[$ | 0 |
| $[190, 200[$ | 1 |

On dit qu'on a regroupé les caractères par classes.
On représente généralement ces résultats sous la forme d'un histogramme.




**Exemple 7.** Reprenons l'exemple des longueurs des feuilles de maïs. Si on réécrit ces données sous forme de classe, on obtient :

| Classe (cm) | Effectif $n_i$ | Fréquence (%) |
|---|---|---|
| $[8, 9[$ | 1 | 5 |
| $[9, 10[$ | 2 | 10 |
| $[10, 11[$ | 3 | 15 |
| $[11, 12[$ | 5 | 25 |
| $[12, 13[$ | 4 | 20 |
| $[13, 14[$ | 3 | 15 |
| $[14, 15[$ | 2 | 10 |
| **Total** | **20** | **100** |



On définit alors comme précédemment les effectifs cumulés croissants et les fréquences cumulées croissantes.

### 14.1.4 Caractéristiques empiriques de position

**Définition 4: Mode**

Le mode d'une série statistique est le caractère (ou la classe de caractères) le plus fréquemment atteint.

**Exemple 8.** • Dans l'exemple des notes dans la classe de 15 élèves ci-dessus, le mode est la note 12. • Dans l'exemple des longueurs des feuilles de maïs, le mode est la longueur 11.
• Dans le cas des tailles, on dit que la classe modale est l'intervalle $[160, 170[$.

**Définition 5: Moyenne**

• La moyenne d'une série statistique $x = (x_1, \dots, x_n)$ est $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$.

• Si $\{x_1, \dots, x_n\} = \{y_1, \dots, y_p\}$ avec $y_1 < y_2 < \cdots < y_p$ d'effectifs respectifs $n_i$ et de fréquences respectives $f_i$, on a

$$\overline{x} = \frac{1}{n}\sum n_i y_i = \sum_{i=1}^{p} f_i y_i.$$

**Exemple 9.** La moyenne des longueurs des feuilles de maïs est :

$$\bar{x} = \frac{\displaystyle\sum_{i=1}^{7} n_i x_i}{n} = \frac{1\times 8 + 2\times 9 + 3\times 10 + 5\times 11 + 4\times 12 + 3\times 13 + 2\times 14}{20}$$

$$\bar{x} = \frac{8 + 18 + 30 + 55 + 48 + 39 + 28}{20} = \frac{226}{20} = 11{,}3 \text{ cm}$$

**Définition 6: Médiane**

Soit $(x_1, \dots, x_n)$ une série statistique avec $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$.
• Si $n = 2p + 1$ est impair, alors on définit la médiane de la série statistique $x$ par la valeur

$$M = x_{p+1}.$$

• Si $n = 2p$ est pair, alors on définit la médiane de la série statistique $x$ par la valeur

$$M = \frac{1}{2}(x_p + x_{p+1}).$$

Autrement dit, la médiane est la valeur du caractère qui partage l'effectif en deux.

**Exemple 10.** • Pour les longueurs des feuilles de maïs, l'effectif total est $n = 20$, donc :

$$\text{médiane} = \frac{10^e \text{ et } 11^e \text{ terme}}{2} = 11$$

$$\boxed{\text{Médiane } M = 11 \text{ cm}}$$

• Reprenons l'exemple des notes du devoir de mathématiques. La classe étant constituée de 15 élèves, la médiane est la 8ème note dans l'ordre croissant. En considérant les effectifs cumulés croissants, on constate que la médiane est $M = 12$.

### 14.1.5 Caractéristiques empiriques de dispersion

**Définition 7: Variance et écart-type**

Soit $x = (x_1, \dots, x_n)$ une série statistique.

1. On définit la variance $s_x^2$ de la série $x$ par
$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$
2. On définit l'écart-type $s_x$ de la série $x$ par
$$s_x = \sqrt{s_x^2}.$$

**Exemple 11.** Pour les longueurs des feuilles de maïs, on a :

$$\sigma^2 = \frac{\sum_{i=1}^{7} n_i(x_i - \bar{x})^2}{n}$$

| $x_i$ | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 2 | 3 | 5 | 4 | 3 | 2 |
| $(x_i - \bar{x})^2$ | 10,89 | 5,29 | 1,69 | 0,09 | 0,49 | 2,89 | 7,29 |

$$\sum_{i=1}^{p} n_i(x_i - \bar{x})^2 = 1(10{,}89) + 2(5{,}29) + 3(1{,}69) + 5(0{,}09) + 4(0{,}49) + 3(2{,}89) + 2(7{,}29) = 56{,}2$$

$$\sigma^2 = \frac{56{,}2}{20} = 2{,}81 \quad \text{et} \quad \sigma = \sqrt{2{,}81} \approx 1{,}68 \text{ cm}$$

**Remarque 2.** • La variance de la série statistique $x$ est en fait la moyenne de la série statistique $(x - \overline{x})^2$.

• On a $s_x^2 = 0 \Leftrightarrow \forall i \in [\![1, n]\!], x_i = \overline{x}$ donc une série statistique est de variance nulle si et seulement si elle est constante égale à sa moyenne.

On retrouve la formule de König-Huyens (que l'on verra en probabilités) :

**Proposition 1: Formule de König-Huygens**

Soit $x = (x_1, \dots, x_n)$ une série statistique.
Alors
$$s_x^2 = \overline{x^2} - \overline{x}^2$$
où $x^2$ est la série statistique $x^2 = (x_1^2, \dots, x_n^2)$.

**Définition 8: Quartiles et déciles**

Soit $x = (x_1, \dots, x_n)$ une série statistique avec $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$.
• Les quartiles de $x$ sont trois valeurs de caractères qui découpent la série $x$ en quatre effectifs égaux : le premier quartile est la valeur séparant le quart constitué des plus faibles valeurs du reste, le deuxième quartile est la médiane, et le troisième quartile est la valeur séparant le quart constitué des plus fortes valeurs du reste.
• De même, les déciles de $x$ sont 9 valeurs découpant la série statistique en 10 effectifs égaux.

**Exemple 12.** • Pour les longueurs des feuilles de maïs, on a :

$$Q_1 = \text{valeur de rang}\frac{n}{4} = 5^e \text{ valeur } \Rightarrow Q_1 = 10 \text{ cm}$$

$$Q_3 = \text{valeur de rang}\frac{3n}{4} = 15^e \text{ valeur } \Rightarrow Q_3 = 12 \text{ cm}$$

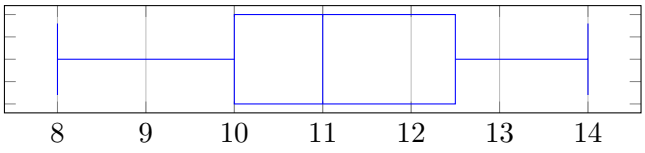
$$\boxed{Q_1 = 10 \text{ cm}, \quad Q_2 = 11 \text{ cm}, \quad Q_3 = 12 \text{ cm}}$$

• Reprenons l'exemple des notes du devoir de mathématiques. La classe est constituée de 15 élèves et $15/4 = 3,75$. On arrondit à l'entier du dessus pour trouver le premier quartile : c'est la 4ème note obtenue dans l'ordre croissant, c'est à dire 10. (Ici, le minimum et le premier quartile sont confondus.)

Le troisième quartile est alors la 12ème note obtenue dans l'ordre croissant, c'est à dire 14.

**Remarque 3.** On représente souvent les quartiles d'une série statistique dans une boîte à moustaches.

Boîte à moustaches horizontale — feuilles de maïs




Longueur (cm)

**Exemple 13.**

$$D_k = \text{valeur du } \frac{kN}{10}^e \text{ individu}$$

$$D_1 : 2^e \text{ individu} \Rightarrow D_1 = 9 \text{ cm}$$
$$D_2 : 4^e \text{ individu} \Rightarrow D_2 = 10 \text{ cm}$$
$$D_3 : 6^e \text{ individu} \Rightarrow D_3 = 10 \text{ cm}$$
$$D_4 : 8^e \text{ individu} \Rightarrow D_4 = 11 \text{ cm}$$
$$D_5 : 10^e \text{ individu} \Rightarrow D_5 = 11 \text{ cm}$$
$$D_6 : 12^e \text{ individu} \Rightarrow D_6 = 12 \text{ cm}$$
$$D_7 : 14^e \text{ individu} \Rightarrow D_7 = 12 \text{ cm}$$
$$D_8 : 16^e \text{ individu} \Rightarrow D_8 = 13 \text{ cm}$$
$$D_9 : 18^e \text{ individu} \Rightarrow D_9 = 13 \text{ cm}$$

## 14.2 Statistique bivariée

Dans cette partie, nous allons étudier la situation suivante. Le mur d'une habitation est constitué par une paroi en béton et une couche de polystyrène d'épaisseur variable $x$ (en cm). On a mesuré, pour une même épaisseur de béton, la résistance thermique $y$ (en $\mathrm{m}^2.^o.W^{-1}$) de ce mur pour différentes valeurs de $x$ et on obtenu le tableau suivant :

| Epaisseur $x_i$ | 2 | 4 | 6 | 8 | 10 | 12 | 15 | 20 |
|---|---|---|---|---|---|---|---|---|
| Résistance $y_i$ | 0,83 | 1,34 | 1,63 | 2,29 | 2,44 | 2,93 | 4,06 | 4,48 |

### 14.2.1 Nuage de points et point moyen

Dans certaines situations, il peut être intéressant d'étudier une série statistique double de taille $n$ portant sur deux caractères quantitatifs $x$ et $y$ (la taille et l'âge, la pression et la température...).

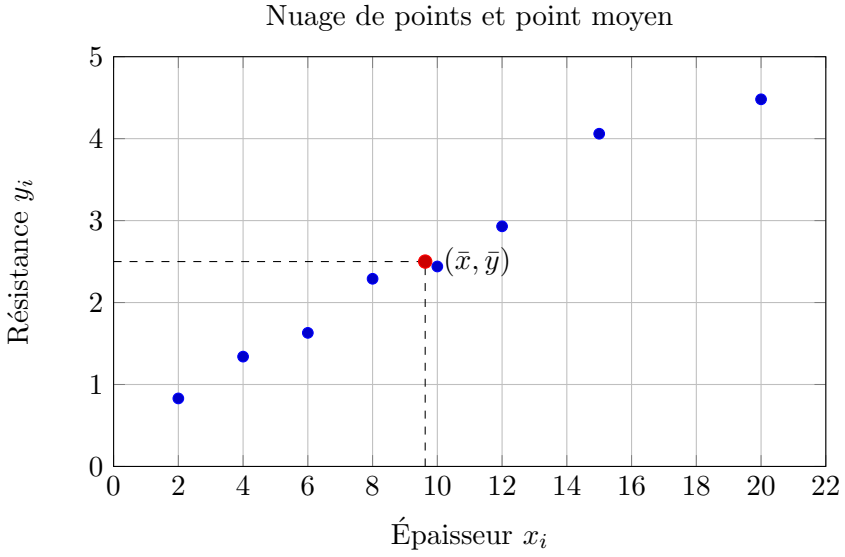
Dans ce cas, on représente les données par un $n$-uplet d'éléments de $\mathbb{R}^2((x_1, y_1), \ldots, (x_n, y_n))$.

On représente graphiquement ces données par un nuage de points de $\mathbb{R}^2$.

**Définition 9: Point moyen d'un nuage de points**

On reprend les notations précédentes.
Soit $(x_i, y_i)_{1\leqslant i\leqslant n}$ une série statistique bivariée.
Soient $\overline{x}$ et $\overline{y}$ les moyennes respectives des séries statistiques $(x_i)_{1\leqslant i\leqslant n}$ et $(y_i)_{1\leqslant i\leqslant n}$.
On appelle point moyen du nuage de points $(x_i, y_i)_{1\leqslant i\leqslant n}$ le point $(\overline{x}, \overline{y})$.

**Exemple 14.** Nuage de points de la résistance en fonction de l'épaisseur :




Les coordonnées du point moyen sont :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{77}{8} = 9{,}625, \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{20}{8} = 2{,}5.$$

$$\boxed{(\bar{x}, \bar{y}) = (9{,}625 \ ; \ 2{,}5)}$$

### 14.2.2 Covariance et coefficient de corrélation

**Définition 10: Covariance**

Soit $(x_i, y_i)_{1\leqslant i\leqslant n}$ une série statistique bivariée.
On appelle covariance de la série $(x, y)$, notée $s_{xy}$, le nombre

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

où $\overline{x}$ et $\overline{y}$ désignent les moyennes respectives des séries statistiques $(x_i)_{1\leqslant i\leqslant n}$ et $(y_i)_{1\leqslant i\leqslant n}$.

**Remarque 4.** • La covariance $s_{xy}$ est en fait la moyenne de la série statistique univariée

$$((x_i - \overline{x})(y_i - \overline{y}))_{1\leqslant i\leqslant n}.$$

• Si $y = x$, on a $s_{xx} = s_x^2$, c'est à dire que la covariance de la série $(x, x)$ est égale à la variance de $x$.

**Proposition 2**

Soit $(x_i, y_i)_{1\leqslant i\leqslant n}$ une série statistique bivariée.
Alors

$$s_{xy} = \overline{xy} - \overline{x} \times \overline{y}$$

où $\overline{xy}$ désigne la moyenne de la série statistique $(x_i y_i)_{1\leqslant i\leqslant n}$.

**Démonstration.** On a

$$\begin{aligned} s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - \overline{x}y_i + \overline{x}\times\overline{y}) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{y}\frac{1}{n}\sum_{i=1}^{n} x_i - \overline{x}\frac{1}{n}\sum_{i=1}^{n} y_i + \frac{1}{n}\sum_{i=1}^{n}\overline{x}\times\overline{y} \\ &= \overline{xy} - \overline{y}\times\overline{x} - \overline{x}\times\overline{y} + \overline{x}\times\overline{y} \\ &= \overline{xy} - \overline{x}\times\overline{y}. \end{aligned}$$

■

**Exemple 15.** La covariance de la série étudiée est :

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y} = \frac{245{,}18}{8} - 9{,}625 \times 2{,}5 = 30{,}6475 - 24{,}0625 = 6{,}585.$$

**Définition 11: Coefficient de corrélation**

Soit $(x_i, y_i)_{1\leqslant i\leqslant n}$ une série statistique bivariée.
On suppose que les séries statistiques $x = (x_i)_{1\leqslant i\leqslant n}$ et $y = (y_i)_{1\leqslant i\leqslant n}$ sont d'écart-types $s_x$ et $s_y$ non nuls.
On définit alors le coefficient de corrélation de la série $(x, y)$ par

$$r_{xy} = \frac{s_{xy}}{s_x s_y},$$

i.e. la covariance de $(x, y)$ divisée par le produit des écart-types de $x$ et $y$.

**Remarque 5.** • On a toujours $r_{xy} \in [-1,1]$.

En effet, $r_{xy} = \dfrac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$ et l'inégalité $|r_{xy}| \leqslant 1$ découle de l'inégalité de Cauchy-Schwarz.

• Plus le coefficient de régression linéaire est proche de 1 en valeur absolue, meilleur est l'ajustement linéaire.

• Lorsque $r = \pm 1$, la droite de régression passe par tous les points du nuage, qui sont donc alignés.

**Exemple 16.** Le coefficient de corrélation de la série étudiée est :

$$r_{xy} = \frac{s_{xy}}{s_x\, s_y} = \frac{6{,}585}{(5{,}5664)(1{,}1978)} \approx 0{,}9876.$$

**Exemple 17.** Soit $x = (x_1, \dots, x_n)$ une série statistique. Soit $(a,b) \in \mathbb{R}^* \times \mathbb{R}$. On considère la série statistique $y = ax + b = (ax_1 + b, \dots, ax_n + b)$.

On a $s_{xy} = \overline{xy} - \overline{x} \times \overline{y} = \overline{x(ax+b)} - \overline{x} \times \overline{ax+b} = a\overline{x^2} + b\overline{x} - a\overline{x}^2 - b\overline{x} = a(\overline{x^2} - \overline{x}^2) = as_x^2$.

D'autre part,

$$s_y^2 = \overline{(ax+b)^2} - \overline{ax+b}^2 = a^2\overline{x^2} + 2ab\overline{x} + b^2 - (a\overline{x}+b)^2 = a^2\overline{x^2} + 2ab\overline{x} + b^2 - a^2\overline{x}^2 - 2ab\overline{x} - b^2$$

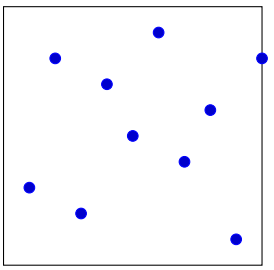
d'où $s_y^2 = a^2 s_x^2$ puis $s_y = |a| s_x$.

Ainsi, $r_{xy} = \dfrac{s_{xy}}{s_x s_y} = \dfrac{as_x^2}{|a|s_x^2} = \dfrac{a}{|a|} = \begin{cases} 1 & \text{si } a > 0 \\ -1 & \text{si } a < 0 \end{cases}$

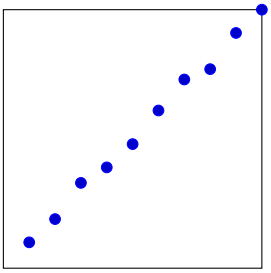
### 14.2.3 Ajustement affine selon la méthode des moindres carrés

Soit $(x_i, y_i)_{1 \leqslant i \leqslant n}$ une série statistique bivariée. On suppose qu'il existe un couple $(i,j) \in [\![1,n]\!]^2$ tel que $x_i \neq x_j$, c'est à dire que les points ne sont pas tous alignés sur une droite verticale.

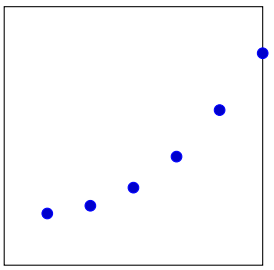
Le nuage de points est un indicateur qui peut s'avérer utile pour vérifier une corrélation entre les caractères.



**Exemple 1**



**Exemple 2**



**Exemple 3**

➢ Si les points sont sous la forme d'un nuage, on peut penser que $x$ et $y$ ne sont pas corrélés (exemple 1).

➢ S'ils semblent dessiner une courbe (exemples 2 et 3), on cherchera à déterminer la nature de la courbe en procédant à un ajustement.

➢ Lorsque les points semblent alignés (exemple 2), on cherche alors à déterminer une droite qui ajuste au mieux ce nuage de points. On dit alors qu'on effectue un ajustement linéaire.

Dans ce qui suit, on se propose de trouver la droite d'équation $y = ax + b$ qui approche au mieux le nuage de points $(x_i, y_i)_{1 \leqslant i \leqslant n}$.

Nous souhaitons pour cela trouver les réels $(a, b) \in \mathbb{R}^2$ qui minimisent la somme

$$\sum_{i=1}^{n} (y_i - (ax_i + b))^2.$$

C'est la somme des carrés des distances de chaque point du nuage de points au point de la droite de même abscisse.

On admet l'existence d'un tel couple $(a, b)$. La droite d'équation $y = ax + b$ est alors appelée droite de régression linéaire de la série statistique $(x, y)$.

**Proposition 3: Régression linéaire selon la méthode des moindres carrés**

La droite de régression linéaire a pour équation $y = ax + b$ avec $a = \dfrac{s_{xy}}{s_x^2}$ et $b = \overline{y} - a\overline{x}$.

**Exemple 18.** Pour notre exemple de la résistance en fonction de l'épaisseur :

$$a = \frac{s_{xy}}{s_x^2} \quad ; \quad b = \bar{y} - a\bar{x}$$

$$\bar{x} = 9{,}625, \quad \bar{y} = 2{,}5$$

$$s_{xy} = 6{,}585, \quad s_x^2 = 30{,}97$$

$$a = \frac{6{,}585}{30{,}97} = 0{,}213$$

$$b = 2{,}5 - 0{,}213 \times 9{,}625 = 0{,}45$$

$$\boxed{y = 0{,}213x + 0{,}45}$$

Nuage de points et droite des moindres carrés