1 Série statistique à une variable

Dans toute cette partie, on considèrera les 3 séries statistiques suivantes :
Série A :
Notes obtenues à un contrôle dans une classe de 30 élèves :

$$2-3-3-4-5-6-6-7-7-7-8-8-8-8-8-9-9-9-9-9-9-10-10-11-11-11-13-13-15-16$$

Série B :
Salaires en euros des employés d'une entreprise :

Salaires	$[900;1200]$	$[1200;1400]$	$[1400;1600]$	$[1600;1800]$	$[1800;2000]$	$[2000;2400]$	**TOTAL**
Effectif	30	30	60	80	40	40	280

Série C :
Proportion d'adhérents à un club sportif dans différentes sections :

- 17% jouent au handball,
- 25% houent au rugby,
- 58% jouent au tennis.

1.1 Moyenne

Définition :
Si une série statistique x_1 ; ... ; x_p prend n_i fois la valeur x_i, pour $i \in \{1, ..., p\}$, alors on a la moyenne pondérée qui vaut

$$\overline{x} = \frac{\qquad\qquad}{\qquad\qquad}$$

- L'évènement "A et B" , noté A∩B, est constitué des issues réalisant à la fois A et B.
- L'évènement "A ou B" , noté A∪B, est constitué des issues réalisant A ou B.
- L'évènement contraire de A, noté $\overline{A}$, est constitué des issues de réalisant pas A.

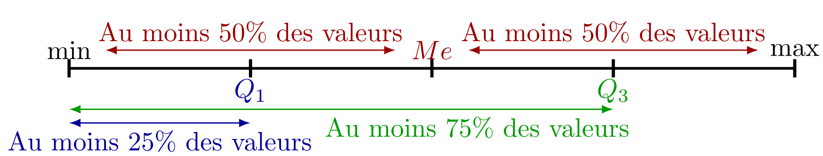
Exemple:

- Dans la **série A**, la moyenne du contrôle est égale à $\overline{m} = \dfrac{254}{30} \approx 8,47$.
- Dans la **série B**, une estimation du salaire moyen est donné par : $\overline{S} = \dfrac{460500}{280} \approx 1644,64$.

1.2 Médiane et quartiles

Définition:

- On appelle médiane Me d'une série statistique à n éléments, la plus petite valeur de la série qui est supérieure ou égale à 50% des valeurs de la série.
  - Si n est impair alors la médiane est la valeur centrale.
  - Si n est pair alors la médiane est la moyenne des deux valeurs centrales.
- On appelle 1° quartile Q_1 (respectivement 3° quatile Q_3) la plus petite valeur de la série qui est supérieure ou égale à 25% (respectivement 75%) des valeurs de la série.



1.3 Ecart interquartiles

Définition:
On appelle écart interquartile la donnée :

$$EQ =$$

1.4 Variance et écart-type

Définition:
Si une série statistique x_1 ; ... ; x_p prend n_i fois la valeur x_i, pour $i \in \{1, ..., p\}$, alors on définit respectivement la variance et l'écart-type par:

$$V = \frac{n_1(x_1 - \bar{x})^2 + ... + n_p(x_p - \bar{x})^2}{n_1 + ... + n_p} \quad \text{et} \quad \sigma = \sqrt{V}$$

L'écart type représente la distance moyenne entre les valeurs de la série et la moyenne.
Plus l'écart-type est petit, ...

2 Série statistique à deux variables

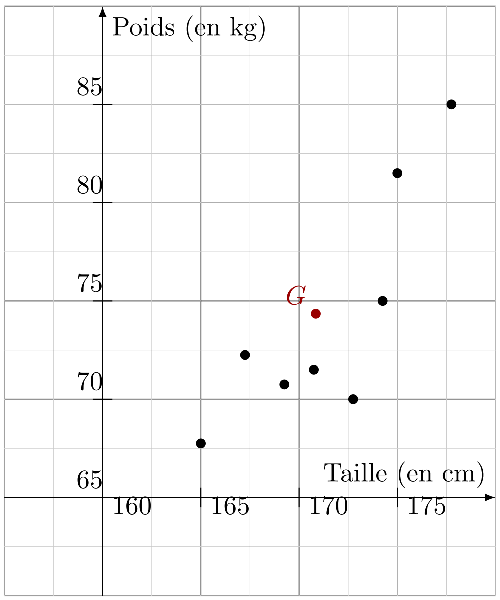
On interroge 8 personnes en leur demandant chacune leur taille et leur poids. On a donc une série statistique x relative à la taille et une série statistique y relative au poids.

Personne	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Taille	165	167	169	171	173	174	175	178
Poids	68	73	71	72	70	75	82	85

Définition:
A une série statistique on associe :

- Un nuage de points, qui est l'ensemble des n points $M_i(x_i; y_i)$ dans un repère du plan.
- Un point moyen dans le même repère : $G(\overline{x}; \overline{y})$ où $\overline{x}$ et $\overline{y}$ sont respectivements les moyennes des séries x et y.

Exemple:
En calculant les moyennes des séries statistiques taille (x) et poids (y), on obtient respectivement $\bar{x} = 171,5$ et $\bar{y} = 74,5$.
On a donc le point moyen $G(171,5; 74,5)$.



Le principe de l'ajustement est de chercher un lien éventuel et simple entre x et y.
Dans le cadre d'un ajustement affine, on cherche à lier x et y par une relation de la forme $y = ax + b$. On obtient alors une droite d'ajustement censée représenter le nuage de points.

- ***Méthode 1:*** On suppose que la droite passe par deux points G et $A(159; 60,5)$.
  On a alors son coefficient directeur $m =$
  On a son ordonnée à l'origine donné par :
  $y_G = \qquad \iff b =$

- ***Méthode 2:*** On suppose que la droite passe par G en connaissant son ordonnée à l'origine.
  On a $y_G = ax_G - 117,58 \iff a \simeq 1,12$.

- ***Méthode 3:*** On suppose que la droite passe par G en connaissant son coefficient directeur.
  On a $y_G = 1,12x_G + b \iff b \simeq -117,58$.