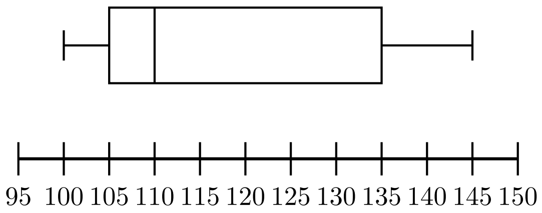
# Statistiques à 1 variable

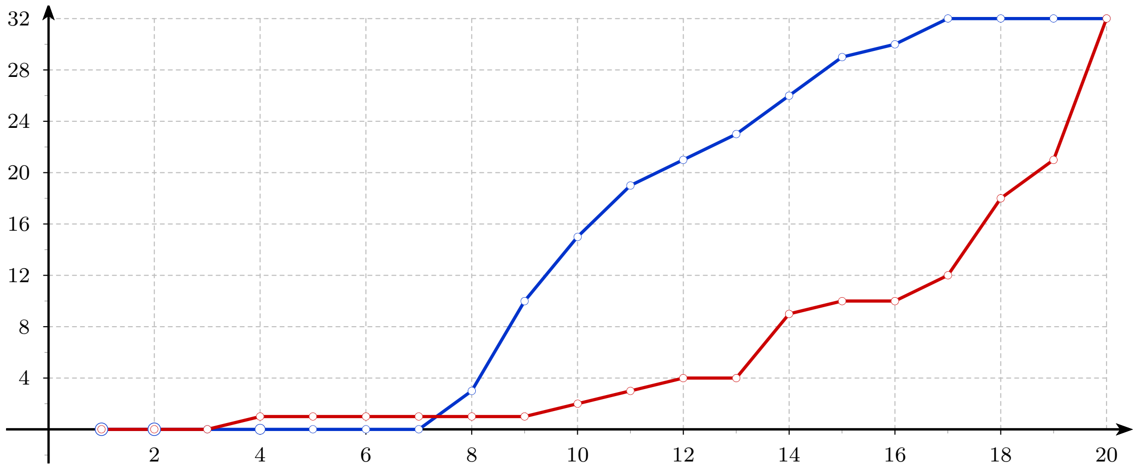
**Exercice n° 1**

Déterminer l'intervalle et l'écart interquartile de la série représentée par le diagramme en boîte suivant.



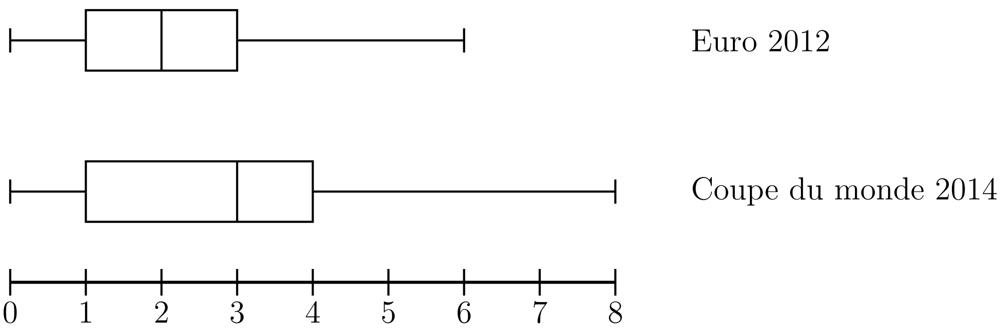

**Exercice n° 2**

Ci-dessous les diagrammes des effectifs cumulés croissants (ECC) pour les notes du dernier DS et IE. Déterminer médiane et quartiles pour chaque épreuve et tracer sur le même graphique les diagrammes en boîtes correspondants.




**Exercice n° 3**

On donne, sur le même graphique, le diagramme en boîte correspondant à la série du nombre de buts marqués par match lors de l'Euro 2012.




1. Lire le minimum, le maximum, la médiane et les premier et troisième quartiles de la série du nombre de buts marqués par match lors de l'Euro 2012 sur le diagramme en boîte correspondant.
2. L'Euro 2012 a-t-il été plutôt plus ou moins *offensif* que la Coupe du monde 2014 ? Argumenter.

# Statistiques à 1 variable

**Exercice n° 4**

Déterminer l'écart-type des séries statistique suivantes à la main et vérifier à l'aide de la calculatrice.

Série 1 :

2 ; 4 ; 8 ; 16 ; 32 ; 64 ; 128 ; 128.

Série 2 :

| Valeur | 6 | 7 | 8 | 9 | 10 | 12 |
|---|---|---|---|---|---|---|
| Effectif | 9 | 2 | 1 | 8 | 7 | 6 |

**Exercice n° 5**

Simplifier $\displaystyle\sum_{i=1}^{5} x_i - \sum_{i=2}^{3} x_i$ puis $\displaystyle\sum_{i=7}^{12} n_i - \sum_{k=6}^{9} n_k$.

**Exercice n° 6** — **Avec des sommes**

**1.** Écrire $\displaystyle\sum_{i=1}^{n}(x_i + y_i)$ sous forme de deux sommes.

**2.** Soit $a$ un réel.

**a.** Simplifier $\displaystyle\sum_{i=1}^{n} a$.

**b.** Écrire $\displaystyle\sum_{i=1}^{n}(x_i + a)$ puis $\displaystyle\sum_{i=1}^{n} a x_i$ en fonction de $\displaystyle\sum_{i=1}^{n} x_i$.

**Exercice n° 7** — **Questions de linéarité**

**1. a.** Quand on ajoute un même nombre $a$ à toutes les valeurs d'une série statistique, que devient la moyenne ? Justifier.

**b.** Quand on ajoute un même nombre $a$ à toutes les valeurs d'une série statistique, que devient l'écart-type ? Justifier.

**2. a.** Quand on multiplie toutes les valeurs d'une série statistique par un même nombre positif $b$, que devient la moyenne ? Justifier.

**b.** Quand on multiplie toutes les valeurs d'une série statistique par un même nombre positif $b$, que devient l'écart-type ? Justifier.

**3. a.** Pour demain, le professeur de Jérémy l'a mis au défi de trouver une série de valeurs de moyenne 13 et d'écart-type 2.

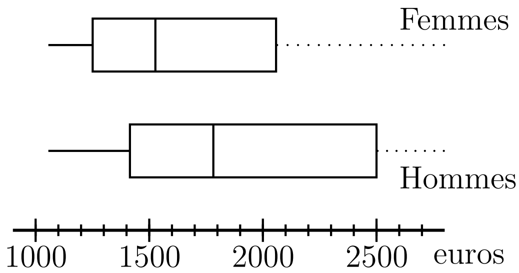
Jérémy a choisi des valeurs au hasard et a obtenu, sur cette série, une moyenne de 12 et un écart-type de 3.

Il décide de multiplier et/ou ajouter un même nombre à toutes ces valeurs afin de répondre au problème posé. Comment s'y prend-il ?

**b.** Les valeurs après modification sont-elles forcément supérieures à celles qui leur correspondent avant modification ? Si non, précisez pour lesquelles ce n'est pas le cas.

**Exercice nº 8** — **Inégalité hommes-femmes**

On représente ci-dessous les diagrammes en boîte des séries des salaires nets mensuels des femmes et des hommes en France en 2010 (les salaires maximaux étant trop élevés, ils n'apparaissent pas sur le graphique).

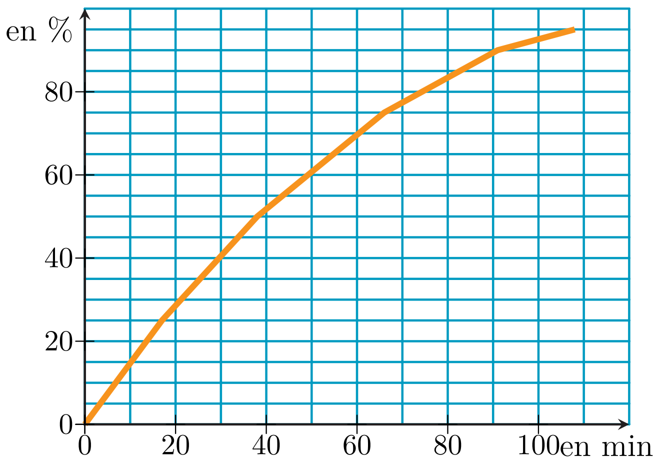



1. **1.** Lire, avec la précision permise par le graphique, le couple médiane-écart interquartile pour les deux séries.
2. **2.** Interpréter les résultats de la question précédente.
3. **3.** Dire si les affirmations suivantes sont vraies ou fausses :
   - la différence entre les salaires médians d'un homme et d'une femme est d'environ 100 € ;
   - plus de 75% des femmes gagnent moins de 2100 € par mois.
4. **4. a.** Calculer la différence de salaire entre une femme et un homme ayant chacun un salaire égal au troisième quartile de leurs séries respectives.

   **b.** En pourcentage, combien cet homme gagne-t-il de plus que cette femme ?

**Exercice nº 9**

On considère le diagramme des fréquences cumulées croissantes de la série des temps d'accès des français à un service de chirurgie cardiaque.




(*Source : Irdes*)

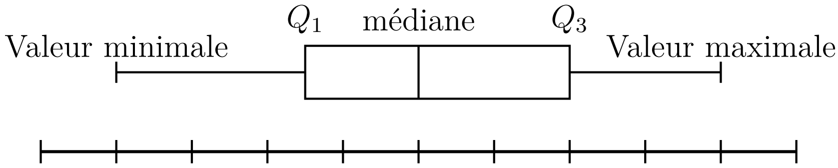
1. **1.** Quel pourcentage de la population a accès à un service de chirurgie cardiaque en moins de 30 minutes ?
2. **2.** Lire les quartiles et la médiane de la série.
3. **3.** Construire le diagramme en boîte de cette série (le maximum n'apparaît pas mais il est de 9 h et 54 min, on ne pourra donc pas le faire apparaître sur le diagramme en boîte).
4. **4.** Les affirmations suivantes sont-elles vraies ou fausses ?

   **a.** Environ 75% des français ont accès à un service de chirurgie cardiaque en environ 66 min ou plus.

   **b.** Il y a environ 50% d'écart entre le temps d'accès des 25% les plus proches et des 25% les plus éloignés d'un service de chirurgie cardiaque.

# Statistiques à 1 variable

**Définition 1** (Diagramme en boîte)**.** Le *diagramme en boîte* d'une série statistique est le graphique suivant :




où l'axe est gradué régulièrement de sorte que l'on puisse y faire figurer les valeurs minimale et maximale, le 1er quartile $Q_1$, la médiane et le 3e quartile $Q_3$.

L'*écart interquartile* est $Q_3 - Q_1$. L'écart interquartile est un **indicateur de dispersion** de la série : plus il est faible, plus la série est homogène.

**Propriété 1** (Linéarité de la moyenne)**.** *Soit $(x_i)$ une série statistique de moyenne $\overline{x}$ et deux réels $a$ et $b$. La série des $(ax_i + b)$ a pour moyenne $a\overline{x} + b$.*

Autrement dit si modifie tous les termes d'une série en leur ajoutant un même nombre (ou en les multipliant tous par un même nombre), la moyenne change en subissant la même transformation.

**Définition 2** (Variance et écart-type)**.** La **moyenne des carrés des écarts à la moyenne** est appelée *variance* d'une série statistiques. On a donc les formules suivantes :

- Soit $x_1$, $x_2$, ..., $x_n$ une série statistique de moyenne $\overline{x}$.
  La *variance* $V$ est donnée par la formule :

$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n} = \frac{\displaystyle\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}.$$

- Si l'on peut écrire la série sous forme de tableau d'effectifs :

| Valeur | $x_1$ | $x_2$ | ... | $x_p$ |
|---|---|---|---|---|
| Effectif | $n_1$ | $n_2$ | ... | $n_p$ |

  La formule précédente de la *variance* devient :

$$V = \frac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + \ldots + n_p (x_p - \overline{x})^2}{n_1 + n_2 + \ldots + n_p} = \frac{\displaystyle\sum_{i=1}^{p} n_i (x_i - \overline{x})^2}{\displaystyle\sum_{i=1}^{p} n_i}.$$

- L'*écart-type* $\sigma$ d'une série statistique est $\sigma = \sqrt{V}$.

**Exemple :** Déterminer la variance et l'écart-type $\sigma$ de la série de valeurs 2 ; 3 ; 4 ; 8 et 13.

On commence par calculer la moyenne $\bar{x} = \dfrac{2+3+4+8+13}{5} = \dfrac{30}{5} = 6$.

On en déduit que la variance est :

$V = \dfrac{(2-6)^2 + (3-6)^2 + (4-6)^2 + (8-6)^2 + (13-6)^2}{5} = \dfrac{16+9+4+4+49}{5}$

Ainsi $V = \dfrac{82}{5} = 16,4$ et l'écart-type est $\sigma = \sqrt{16,4} \simeq 4,05$.

**Remarques**

- La variance et l'écart-type sont des indicateurs de dispersion autour de la moyenne, plus ils sont petits, plus la série est homogène.
- Généralement, on détermine la variance et l'écart-type à l'aide de la calculatrice.
- La moyenne $\overline{x}$ et l'écart-type $\sigma$ s'expriment dans la même unité que les valeurs de la série. Cela a un donc sens de parler des intervalles $[\overline{x} - \sigma ; \overline{x} + \sigma]$, $[\overline{x} - 2\sigma ; \overline{x} + 2\sigma]$, etc. qui sont souvent utilisés en statistiques. Dans l'exemple précédent, toutes les valeurs sauf 13 sont dans l'intervalle $[\overline{x} - \sigma ; \overline{x} + \sigma]$.