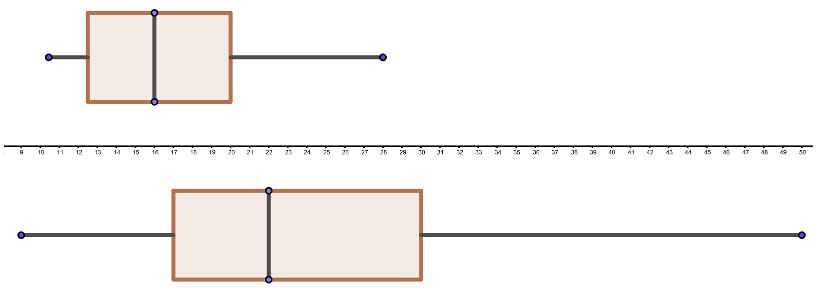
Statistiques - corrections

▷ **Exercice 2.** On donne dans le tableau suivant, les salaires annuels en milliers d'euros en Ile-de-France sur un groupe de 1000 personnes :

salaire	[9 ;10[	[10 ;11[	[11 ;15[	[15 ;20[	[20 ;25[	[25 ;30[	[30 ;40[	[40 ;50]
centres x_i	9,5	10,5	13	17,5	22,5	27,5	35	45
effectifs n_i	50	50	50	256	244	125	125	100
ECC $n_i \nearrow$	50	100	150	406	650	775	900	1000

1. Calculer le salaire annuel moyen en Ile-de-France en indiquant les calculs effectués. (arrondir au millier d'euros)
$\overline{x} = \dfrac{1}{\mathrm{N}}\sum n_i x_i = \dfrac{50\times 9,5 + 50\times 10,5 + \cdot + 100\times 45}{1000} = \dfrac{23932,5}{1000} = 23,9325 \approx 24$
2. Compléter la troisième ligne du tableau (effectifs cumulés croissants).
3. Pour la répartition des salaires en Ile-de-France, on donne $\mathrm{Q}_1 = 17$, $m_e = 22$ et $\mathrm{Q}_3 = 30$.
Le diagramme en boîte ci-dessous correspond aux salaires annuels en milliers d'euros en province (pour les régions françaises hors Ile-de-France).
Compléter avec le diagramme en boîte pour les salaires de la régions Ile-de-France.



4. Donner l'écart inter-quartile pour chacune des deux séries de données.
En observant ces deux diagrammes, que peut-on dire des salaires en Ile de France et en province ?
Pour les salaires en province, l'écart inter-quartile est $20 - 12,5 = 7,5$ et en pour les salaires en Ile-de-France, il est égal à $30 - 17 = 13$.
D'après les box-plots, on constate que les salaires en Ile-de-France sont plus étendus et plus dispersés qu'en province.

▷ **Exercice 3.** Le tableau ci-dessous fournit la répartition des notes de trois classes de trente élèves à un devoir commun :

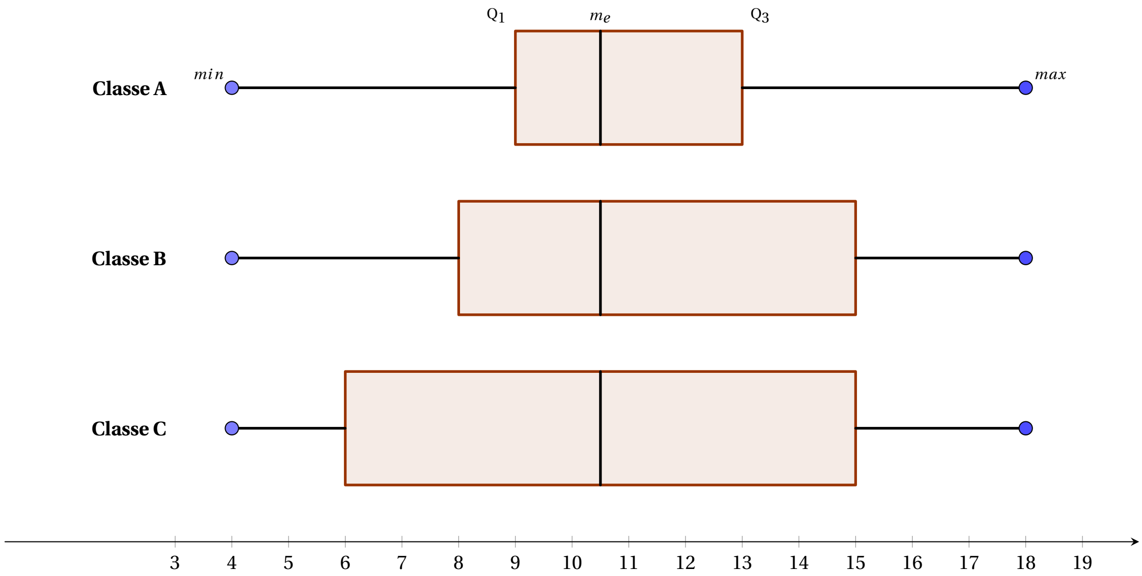
notes	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
classe A	1	0	0	1	3	4	6	2	5	3	2	0	2	0	1
classe B	2	1	2	2	1	2	5	3	2	1	1	3	1	2	2
classe C	3	4	1	2	0	0	5	1	0	0	4	4	3	1	2

Calculer la moyenne, l'écart-type, la médiane, les premier et troisième quartiles, l'écart interquartile de chacune de ces séries. Réaliser les box-plot puis commenter les résultats.

Classe A : $\begin{cases} \overline{x} = 11 \\ \sigma \approx 2,88 \\ \mathrm{Q}_1 = 9 \\ m_e = 10,5 \\ \mathrm{Q}_3 = 13 \\ \mathrm{Q}_3 - \mathrm{Q}_1 = 4 \end{cases}$

Classe B : $\begin{cases} \overline{x} = 11 \\ \sigma \approx 4,07 \\ \mathrm{Q}_1 = 8 \\ m_e = 10,5 \\ \mathrm{Q}_3 = 15 \\ \mathrm{Q}_3 - \mathrm{Q}_1 = 7 \end{cases}$

Classe C : $\begin{cases} \overline{x} = 11 \\ \sigma \approx 4,70 \\ \mathrm{Q}_1 = 6 \\ m_e = 10,5 \\ \mathrm{Q}_3 = 15 \\ \mathrm{Q}_3 - \mathrm{Q}_1 = 9 \end{cases}$



On remarque que les notes de la classe C sont les plus dispersées autour de la médiane (écart inter-quartile supérieur aux deux autre classes). La plus grande dispersion est confirmée par l'écart-type qui est le plus élevé pour la classe C.

Déterminer la proportion d'élèves de chaque classe se trouvant dans l'intervalle $[\,\overline{x}-\sigma\,;\,\overline{x}+\sigma\,]$:

Classe A : $[\,\overline{x}-\sigma\,;\,\overline{x}+\sigma\,] = [\,8,12\,;\,13,88$. Cet intervalle contient les notes comprises entre 9 et 13, soit 20 notes pour la classe A, c'est à dire $\dfrac{20}{30} \approx 67\%$ des notes de la classe A.

Classe B : $[\,\overline{x}-\sigma\,;\,\overline{x}+\sigma\,] = [\,6,93\,;\,15,07$. Cet intervalle contient les notes comprises entre 6 et 15, soit 22 notes pour la classe B, c'est à dire $\dfrac{22}{30} \approx 73\%$ des notes de la classe B.

Classe C : $[\,\overline{x}-\sigma\,;\,\overline{x}+\sigma\,] = [\,6,3\,;\,15,7$. Cet intervalle contient les notes comprises entre 6 et 15, soit 20 notes pour la classe C, c'est à dire $\dfrac{20}{30} \approx 67\%$ des notes de la classe C.

▷ **Exercice 4.** On a relevé dans une entreprise le temps en minutes consacré à la pratique d'un sport par semaine. Il s'agit d'une série statistique à caractère continu. On obtient le tableau ci-dessous :

Temps en minutes : x_i	[0;20[	[20;40[	[40;60[	[60;100[	[100;140[	[140;200[
Centre des classes :	10	30	50	80	120	170
Effectifs : n_i	29	43	47	12	5	2
Effectifs cumulés croissants :	29	72	119	131	136	138
Fréquences en % : f_i	21,01	31,16	34,06	8,70	3,62	1,45
Fréquences cumulées croissantes :	21,01	52,17	86,23	84,93	98,55	100

1. Compléter les lignes du tableau.
2. Déterminer la classe médiane (intervalle auquel appartient la médiane).

   $m_e \in [\,20\,;\,40\,]$
3. Calculer la moyenne et l'écart-type de cette série.

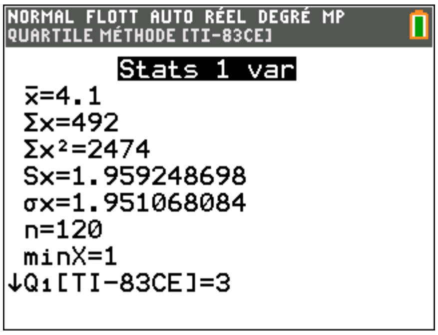
   En utilisant le centre des classes, on obtient : $\overline{x} = \dfrac{1}{N}\sum n_i x_i = \dfrac{5830}{138} \approx 42\,min$

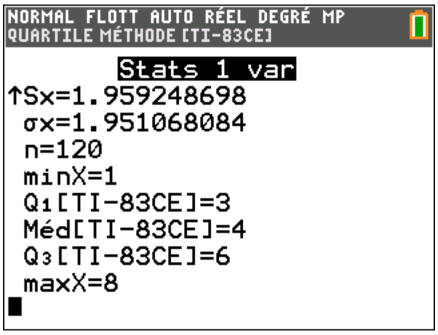
   $V = \dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2 = \dfrac{365700}{138} - \left(\dfrac{5830}{138}\right)^2 \approx 865,24$ donc $\sigma = \sqrt{V} \approx 29$

▷ **Exercice 5.** Une étude sur le nombre d'employés dans les commerces du centre d'une petite ville a donné les résultats suivants :

Nombre d'employés	1	2	3	4	5	6	7	8
Effectif	11	18	20	24	16	14	11	6

Déterminer la médiane, la moyenne, les 1$^{\text{er}}$ et 3$^{\text{ème}}$ quartiles, la variance et l'écart-type de cette série statistique. Construire la box-plot.





On obtient : $\overline{x} = \dfrac{492}{120} = 4,1$, $V = \dfrac{2474}{120} - 4,1^2 \approx 3,81$, $\sigma \approx 1,95$, $Q_1 = 3$, $m_e = 4$, $Q_3 = 6$