Statistiques - corrections

▷ **Exercice 1.** On a demandé à 50 personnes prenant le bus, le nombre de fois où elles ont utilisé ce type de transport pendant la semaine écoulée. Voici les résultats :

Nombre de voyages en bus : x_i	1	2	3	4	5	6	7	8	9	10
Effectifs : n_i	3	3	5	7	6	9	5	4	5	3
Effectifs cumulés croissants : $n_i \nearrow$	3	6	11	18	24	33	38	42	47	50
Fréquences en % : f_i	6%	6%	10%	14%	12%	18%	10%	8%	10%	6%
Fréquences cumulées croissantes : $f_i \nearrow$	6%	12%	22%	36%	48%	66%	76%	84%	94%	100%

1. Tableau.
2. Déterminer la médiane et la moyenne.

   L'effectif total est N = 50 et il est pair. $\frac{N}{2} = 25$ donc la médiane se situe entre la 25$^{\text{ième}}$ valeur qui est égale 6 et la 26$^{\text{ième}}$ valeur qui vaut également 6 donc $m_e = 6$.

   $\overline{x} = \dfrac{3\times 1 + 3\times 2 + \cdots + 3\times 10}{50} = \dfrac{278}{50} = 5,56$
3. Déterminer les premier et troisième quartiles.

   - **1$^{\text{er}}$ quartile** : $\dfrac{25}{100}\times 50 = 12,5$ donc Q_1 est la 13$^{\text{ième}}$ valeur : $Q_1 = 4$
   - **3$^{\text{ème}}$ quartile** : $\dfrac{75}{100}\times 11 = 37,5$ donc Q_3 est la 38$^{\text{ième}}$ valeur : $Q_3 = 7$
4. Calculer la variance et l'écart-type.

   $V = \dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2 = \dfrac{3\times 1^2 + 3\times 2^2 + \cdots + 3\times 10^2}{50} - 5,56^2 = \dfrac{1852}{50} - 5,56^2 = 6,1264$ donc $\sigma = \sqrt{V} \approx 2,48$

▷ **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 :

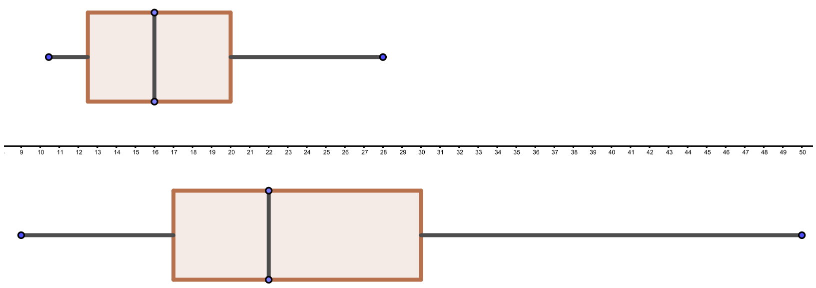
salaire	[9;10[	[10;11[	[11;15[	[15;20[	[20;25[	[25;30[	[30;40[	[40;50]
centres x_i	9,5	10,5	13	17,5	22,5	27,5	35	45
effectifs n_i	50	50	50	256	244	125	125	100
ECC $n_i \nearrow$	50	100	150	406	650	775	900	1000

1. Calculer le salaire annuel moyen en Ile-de-France en indiquant les calculs effectués. (arrondir au millier d'euros)

   $\overline{x} = \dfrac{1}{N}\sum n_i x_i = \dfrac{50\times 9,5 + 50\times 10,5 + \cdots + 100\times 45}{1000} = \dfrac{23932,5}{1000} = 23,9325 \approx 24$
2. Compléter la troisième ligne du tableau (effectifs cumulés croissants).
3. Pour la répartition des salaires en Ile-de-France, on donne $Q_1 = 17$, $m_e = 22$ et $Q_3 = 30$.

   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 $30-17=13$.

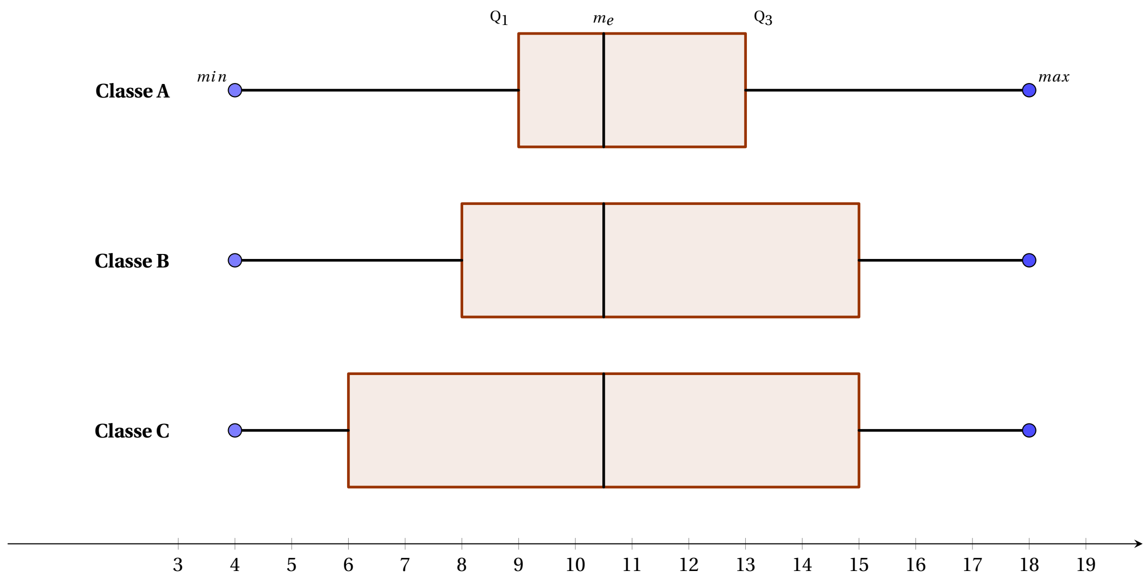
D'après les box-plots, on constate que les salaires en Ile-de-France sont plus étendus et plus dispersés qu'en province.

▷ **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 : $\begin{cases} \overline{x}=11 \\ \sigma \approx 2,88 \\ Q_1 = 9 \\ m_e = 10,5 \\ Q_3 = 13 \\ Q_3 - Q_1 = 4 \end{cases}$ **Classe B :** $\begin{cases} \overline{x}=11 \\ \sigma \approx 4,07 \\ Q_1 = 8 \\ m_e = 10,5 \\ Q_3 = 15 \\ Q_3 - Q_1 = 7 \end{cases}$ **Classe C :** $\begin{cases} \overline{x}=11 \\ \sigma \approx 4,70 \\ Q_1 = 6 \\ m_e = 10,5 \\ Q_3 = 15 \\ Q_3 - Q_1 = 9 \end{cases}$



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.

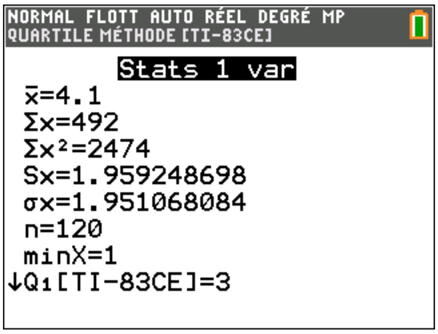
   En utilisant le centre des classes, on obtient : $\overline{x} = \dfrac{1}{N}\sum n_i x_i = \dfrac{5830}{138} \approx 42\,min$

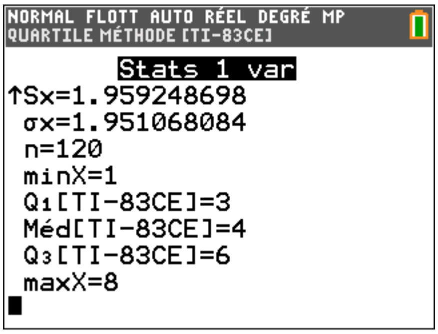
   $V = \dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2 = \dfrac{365700}{138} - \left(\dfrac{5830}{138}\right)^2 \approx 865,24$ donc $\sigma = \sqrt{V} \approx 29$

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

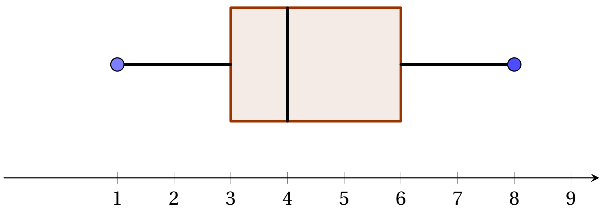
Nombre d'employés	1	2	3	4	5	6	7	8
Effectif	11	18	20	24	16	14	11	6

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





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

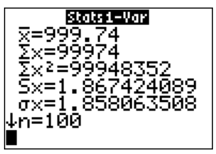


▷ **Exercice 6.** Une conserverie alimentaire fabrique des boîtes de légumes. Afin de vérifier l'état de fonctionnement de la chaîne de remplissage, on a pesé un lot de 100 boîtes de conserves :

Masse en g.	995	996	997	998	999	1000	1001	1002	1003	1004	1005
Nombre de boîtes	3	4	6	7	14	35	20	5	4	1	1
ECC	3	7	13	20	34	69	89	94	98	99	100

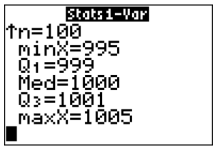
1. Calculer la masse moyenne d'une boîte ainsi que l'écart-type de cette série.

   - $\overline{x} = \dfrac{1}{N}\sum n_i x_i = \dfrac{99974}{100} = 999,74$
   - $V = \dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2 = \dfrac{99948352}{100} - 999,74^2 = 3,4524$ donc $\sigma = \sqrt{V} \approx 1,86$

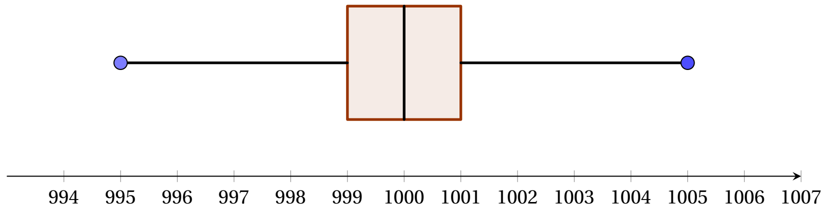


2. Donner la médiane, le premier quartile et le troisième quartile de la série. Construire la box-plot.

Le tableau a été complété par les effectifs cumulés croissants (ECC) :



- L'effectif total est N = 100 et il est pair. $\frac{N}{2} = 50$ donc la médiane se situe entre la 50$^{\text{ième}}$ valeur qui est égale 1000 et la 51$^{\text{ième}}$ valeur qui vaut également 1000 donc $m_e = 1000$.
- **1$^{\text{er}}$ quartile** : $\frac{25}{100} \times 100 = 25$ donc Q_1 est la 25$^{\text{ième}}$ valeur : $Q_1 = 999$
- **3$^{\text{ème}}$ quartile** : $\frac{75}{100} \times 11 = 75$ donc Q_3 est la 75$^{\text{ième}}$ valeur : $Q_3 = 1001$



3. Quel pourcentage des boîtes ont une masse dans l'intervalle $[998; 1002]$?

$7+14+35+20+5 = 81$ boîtes ont une masse dans l'intervalle $[998; 1002]$ soit $\frac{81}{100} = 81\%$ des boîtes.

4. On considère que la chaîne fonctionne correctement si l'écart entre la moyenne et 1000 est inférieur à 0,5 et si le pourcentage de boîtes en dehors de $[998; 1002]$ est strictement inférieur à 20%. La chaîne fonctionne-t-elle correctement ?

$$\left.\begin{array}{l} 1000 - \overline{x} = 1000 - 999,74 = 0,26 < 0,5 \\ 100 - 81 = 19 \text{ donc } 19\% \text{ des boîtes sont en dehors de l'intervalle } [998;1002] \text{ et } 19 < 20 \end{array}\right\} \text{donc la chaîne fonctionne bien.}$$

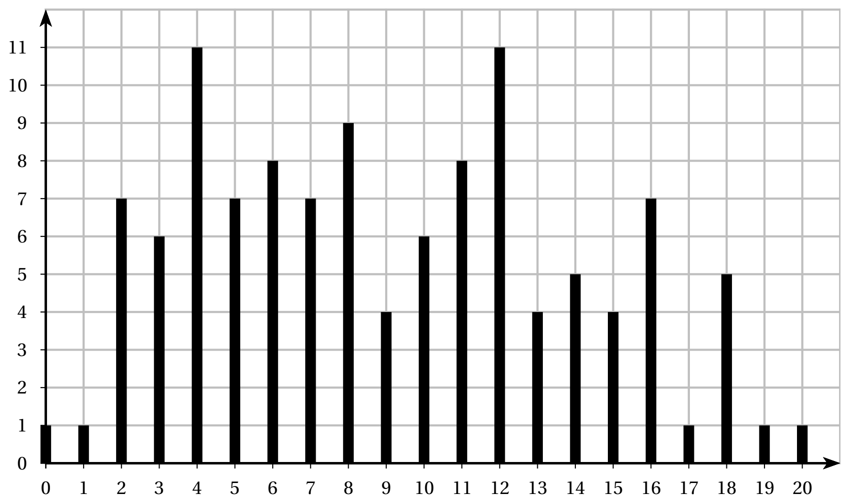
5. Un autre échantillon de 200 boîtes est testé, et sa masse moyenne est de 1000,43g. Quelle est la masse moyenne des 300 boîtes testées ?

$$\overline{X} = \frac{100 \times \overline{x} + 200 \times 1000,43}{300} = \frac{100 \times 999,74 + 200 \times 1000,43}{300} = 1000,2$$

6. L'autre échantillon a 20,5% de boîtes dont le poids est hors-norme. La synthèse des deux tests remet-elle en cause les résultats du premier ?

Dans le deuxième échantillon, il y a $\frac{20,5}{100} \times 200 = 41$ boîtes hors-norme, soit sur les deux échantillons, un total de $41 + 19 = 60$ boîtes, c'est à dire $\frac{60}{300} = 20\%$. La deuxième condition n'est plus vérifiée...

▷ **Exercice 7.** Le diagramme ci-dessous représente les notes obtenues en mathématiques lors du brevet par les élèves d'un collège. La hauteur de chaque barre représente le nombre d'élèves ayant obtenus la note correspondante.

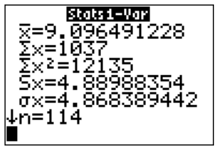


1. Déterminer la moyenne, la médiane, les quartiles et l'écart-type de la série des notes.

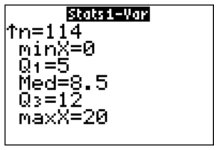
On peut résumer cette situation par le tableau suivant :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectifs	1	1	7	6	11	7	8	7	9	4	6	8	11	4	5	4	7	1	5	1	1
ECC	1	2	9	15	26	33	41	48	57	61	67	75	86	90	95	99	106	107	112	113	114

On en déduit à l'aide de la calculatrice :

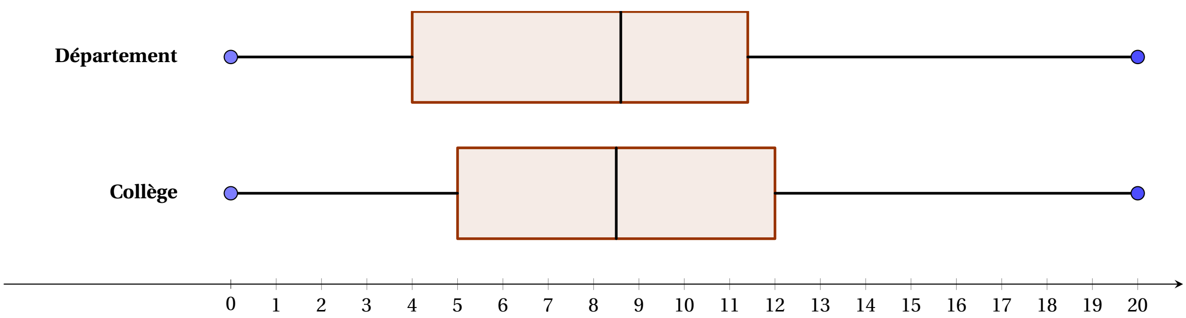


- $\overline{x} = \dfrac{1}{N}\sum n_i x_i = \dfrac{1037}{114} \simeq 9,10$
- $V = \dfrac{1}{N}\sum n_i x_i^2 - \overline{x}^2 = \dfrac{12135}{114} - \left(\dfrac{1037}{114}\right)^2 \simeq 23,70$ donc $\sigma = \sqrt{V} \simeq 4,87$
- L'effectif total est N = 114 et il est pair. $\dfrac{N}{2} = 57$ donc la médiane se situe entre la 57$^{\text{ième}}$ valeur qui est égale 8 et la 58$^{\text{ième}}$ valeur qui est égale à 9 donc $m_e = 8,5$.



- **1$^{\text{er}}$ quartile** : $\dfrac{25}{100} \times 114 = 28,5$ donc Q_1 est la 29$^{\text{ième}}$ valeur : $Q_1 = 5$
- **3$^{\text{ème}}$ quartile** : $\dfrac{75}{100} \times 114 = 85,5$ donc Q_3 est la 86$^{\text{ième}}$ valeur : $Q_3 = 12$

2. Représenter le diagramme en boîte correspondant avec pour échelle 0,5 cm pour un point. Garder la place pour faire un second diagramme au dessus.



3. Au niveau départemental, les résultats obtenus sont donnés ci-dessous.

Notes	[0;4[	[4;8[	[8;12[	[12;16[	[16;20[
Pourcentage des candidats	25	20	35	15	5
Centre des classes	2	6	10	14	18

a) Calculer la moyenne de cette série.

On dispose des fréquences et non des effectifs donc $\overline{x} = \sum f_i.x_i = \dfrac{25}{100}\times 2 + \dfrac{20}{100}\times 6 + \dfrac{35}{100}\times 10 + \dfrac{15}{100}\times 14 + \dfrac{5}{100}\times 18 = 8,2$

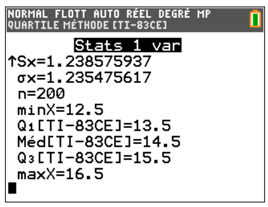
b) Pour cette série, on donne $Q_1 = 4$, $m_e = 8,6$ et $Q_3 = 11,4$ Construire au dessus du diagramme précédent, la box-plot de cette série.

4. Comparer les résultats du collège aux résultats départementaux

On constate que les répartitions des notes (dispersions) sont assez semblables au niveau du collège et du département car les écarts inter-quartiles sont très proches. Néanmoins, même si médianes et moyennes sont peu différentes, la comparaison tourne à l'avantage du collège dont les résultats sont légèrement meilleurs (la boîte est décalée vers la droite).

▷ **Exercice 8.** Une étude sur la durée de vie en heures de 200 ampoules électriques a donné les résultats suivants :

Durée de vie en centaine d'heures	[12; 13[	[13; 14[	[14; 15[	[15; 16[	[16; 17[
Effectif	28	46	65	32	29
Fréquences en %	14	23	32,5	16	14,5
Fréquences cumulées croissantes	14	37	69,5	85,5	100

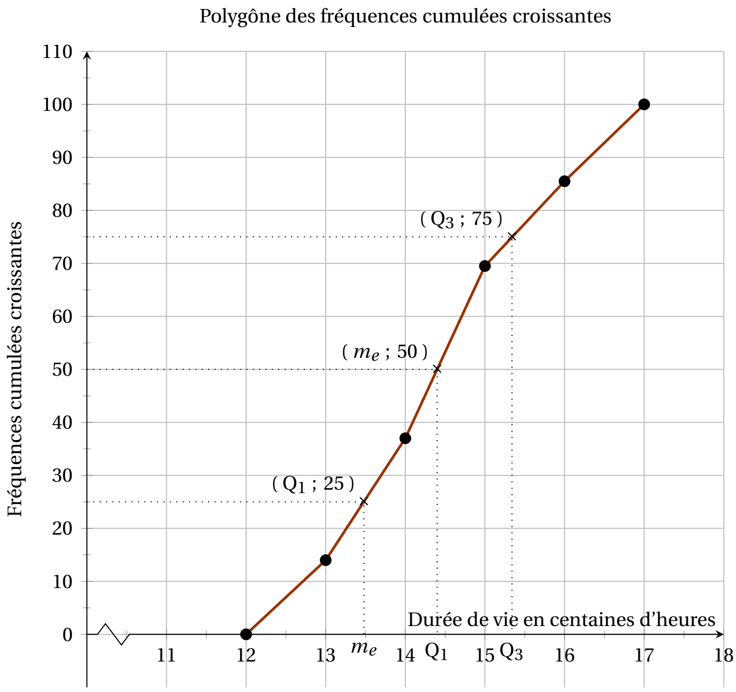


Déterminer la médiane, la moyenne, les 1$^{\text{er}}$ et 3$^{\text{ème}}$ quartiles, la variance et l'écart-type de cette série statistique. Construire la box-plot.
On obtient facilement la moyenne, la variance et l'écart-type de cette série statistique avec le centre des classes et l'utilisation de la calculatrice : $\overline{x} = 14,44$, $V = 1,5264$ et $\sigma = \sqrt{V} \simeq 1,24$

Pour la médiane, et les quartiles, on construit le polygone des fréquences cumulées croissantes. Pour cela, on commence par placer les points dont les coordonnées sont :
$\begin{cases} \text{en abscisse : la borne supérieur de la classe} \\ \text{en ordonnée : la FFC correspondant à cette classe} \end{cases}$
ainsi que le point qui a pour abscisse la borne inférieure de la première classe et 0 en ordonnée. Ces points sont ensuite reliés par des segments de droite.

- La médiane est l'abscisse du point d'ordonnée 50.
- Le premier quartile est l'abscisse du point d'ordonnée 25.
- Le troisième quartile est l'abscisse du point d'ordonnée 75.

Ici, on lit : $\begin{cases} Q_1 \simeq 13,5 \\ m_e \simeq\simeq 14,4 \\ Q_3 \simeq 15,3 \end{cases}$



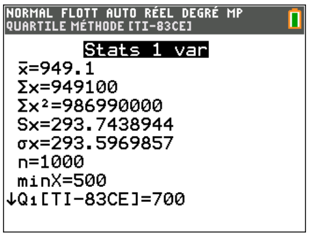
▷ **Exercice 9.** On a testé les durées de vie (en heures) de deux ampoules différentes : 1000 ampoules de type A et 700 du type B.

Durée de vie	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800
Effectifs A	90	97	100	110	130	107	105	100	72	28	33	17	9	2
Effectifs cumulés A	90	187	287	397	527	634	739	839	911	939	972	989	998	1000
Effectifs B	72	70	68	70	69	62	57	53	50	46	40	23	13	8

1. Calculer la durée de vie moyenne et l'écart type pour chaque type d'ampoule, arrondis à l'heure près. Commenter ces résultats.

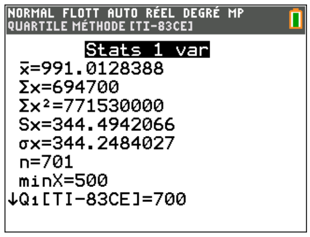
Série A

$\begin{cases} \overline{x}_A = 949,1 \\ \sigma_A \simeq 293,6 \end{cases}$



Série B

$\begin{cases} \overline{x}_B \simeq 991,0 \\ \sigma_B \simeq 344,2 \end{cases}$



On remarque que la durée de vie moyenne des ampoules de type B est supérieure à celle de type A mais que la dispersion des valeurs est également plus importante.

2. Pour chacune des deux séries, déterminer le pourcentage d'ampoules dont la durée de vie a un écart à la moyenne inférieur à l'écart-type.

Ampoules de type A

$100+110+130+107+105+100 = 652$ ampoules de type A ont une durée de vie qui se situe dans l'intervalle

$$[\,\overline{x}_A - \sigma_A \,;\, \overline{x}_A + \sigma_A\,] \approx [\,655,5 \,;\, 1242,7\,]$$

soit 65,2 %

Ampoules de type B

$68+70+69+62+57+53+50 = 429$ ampoules de type B ont une durée de vie qui se situe dans l'intervalle

$$[\,\overline{x}_B - \sigma_B \,;\, \overline{x}_B + \sigma_B\,] \approx [\,646,8 \,;\, 1335,2\,]$$

soit 42,9 %

3. Après avoir complété les effectifs cumulés croissants pour les ampoules de type A, déterminer la médiane et les quartiles en justifiant les résultats ; donner la signification du troisième quartile.

- L'effectif total est N = 1000 et il est pair. $\dfrac{N}{2} = 500$ donc la médiane se situe entre la $500^{\text{ième}}$ valeur qui est égale 900 et la $501^{\text{ième}}$ valeur qui est aussi égale à 900 donc $m_e = 900$.

- **1^{er} quartile** : $\dfrac{25}{100} \times 1000 = 250$ donc Q_1 est la $250^{\text{ième}}$ valeur : $Q_1 = 700$
- **$3^{ème}$ quartile** : $\dfrac{75}{100} \times 1000 = 750$ donc Q_3 est la $750^{\text{ième}}$ valeur : $Q_3 = 1200$

Au moins 75% des ampoules de type A ont une durée de vie inférieure ou égale à $Q_3 = 1200$.

4. Le constructeur des ampoules de type A affirme que “plus de 75% de ses ampoules durent 700 heures ou plus”, Cette affirmation est-elle vraie ?

On sait, puisque $Q_1 = 700$, qu'au moins 25% des valeurs sont inférieures ou égales à 700 donc au moins 75% sont supérieures ou égales à 700.

On constate qu'il y en a $1000 - 90 - 97 = 813$ soit $81,3\%$ des valeurs.

▷ **Exercice 10.** Des sauteurs à la perche ont relevé leurs performances au cours des derniers mois :

1^{er} sauteur

Hauteur	4,70	4,80	4,85	4,90	4,95	5,00	5,05	5,10	5,20
Nombre de sauts	1	1	1	3	12	4	1	1	1

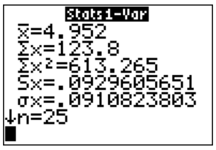
$2^{ème}$ sauteur

Hauteur	4,60	4,70	4,75	4,80	4,85	4,90	4,95	5,00	5,05	5,10	5,15	5,20
Nombre de sauts	3	2	2	3	2	2	1	3	2	1	1	3

Déterminer maintenant la moyenne et l'écart-type de chaque série. Comparer les performances des deux sportifs.

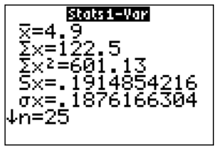
1^{er} sauteur

$$\begin{cases} \overline{x}_1 = 4,952 \\ \sigma_1 \simeq 0,091 \end{cases}$$



$2^{ème}$ sauteur

$$\begin{cases} \overline{x}_2 = 4,9 \\ \sigma_2 \simeq 0,188 \end{cases}$$



On remarque que le $2^{ème}$ sauteur a des performances assez proches en moyenne à celles du 1^{er} sauteur mais beaucoup moins régulière car l'écart-type de sa série est supérieur à celui du 1^{er} sauteur.