CHAPITRE 7

STATISTIQUES UNIVARIÉES

1. Exercices d'application 1
2. En autonomie 5
3. Approfondissement 7

1. Exercices d'application

☼ **Exercice 1** (Salaires dans une entreprise) **:** On a relevé les salaires annuels d'un échantillon de 30 personnes. Les valeurs sont les suivantes en centaines d'euros présentées par ordre croissant.

186	201	211	238	260	277	277	281	292	295
318	325	325	328	336	431	432	496	543	567
571	585	623	625	696	747	788	871	944	959

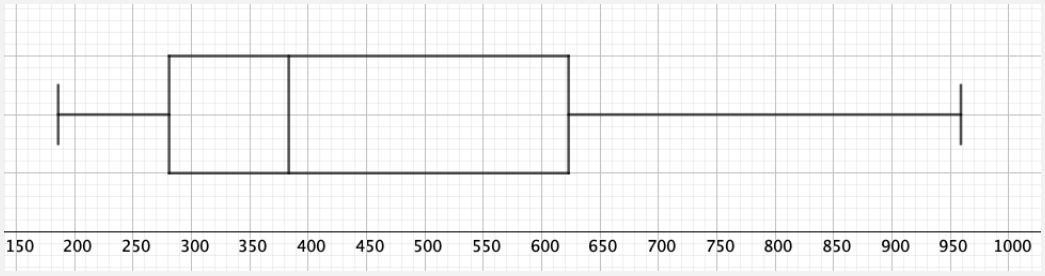
1. Donner la médiane de la série.
2. Déterminer les quartiles et l'intervalle interquartile.
3. Dessiner la boîte à moustaches.
4. Déterminer la moyenne et la variance de la série.
5. Quels sont les salaires moyen et médian?

1. La série est constituée de 30 valeurs, on en déduit que la médiane est située entre la 15^{e} valeur et la 16^{e} valeur :
$$\boxed{M_e = \frac{336+431}{2} = \frac{767}{2} = 383{,}5}$$
2. Pour calculer le premier quartile, on calcule le quart de l'effectif total : $\frac{30}{4} = 7{,}5$
Donc le premier quartile Q_1 est la 8^{e} valeur.
De même, pour calculer le troisième quartile, on calcule les trois quarts de l'effectif total : $\frac{3}{4} \times 30 = 22{,}5$
Donc le troisième quartile Q_3 est la 23^{e} valeur.

Comme les valeurs sont triées en ordre croissant, on trouve facilement

$$Q_1 = 281 \quad \text{et} \quad Q_3 = 623.$$

On en déduit que l'écart interquartiles est $\boxed{Q_3 - Q_1 = 623 - 281 = 342}$.

3. 

4. $$\overline{x} = \frac{14028}{30} = 467{,}6$$

On calcule ensuite la variance à l'aide de la formule de Koenig-Huygens : $s_x^2 = \overline{x^2} - \overline{x}^2$.
D'après le calcul précédent, $\overline{x}^2 = 467{,}7^2 = 218\,649{,}76$.
De plus, en déterminant la série x^2 on obtient

$$\sum_{i=1}^{30} x_i^2 = 186^2 + 201^2 + \cdots + 959^2 = 8\,095\,910$$

On obtient alors

$$\overline{x^2} = \frac{8\,095\,910}{30} \approx 269\,863{,}67$$

On en déduit donc

$$s_x^2 \approx 269\,863{,}67 - 218\,649{,}76 = 51\,213{,}9$$

Ainsi, $\boxed{\text{La moyenne de la série est environ 467,7 et la variance est environ 51 213}}$

5. La série statistique donne les salaires en centaines d'euros. On en déduit donc que $\boxed{\text{le salaire annuel moyen est 46 760 €}}$, et $\boxed{\text{le salaire annuel médian est 38 350 €}}$.

☼ **Exercice 2** (Nombre d'enfants par couple) **:** On considère le tableau suivant donnant le nombre d'enfants par couple dans un échantillon de population de 100 couples.

nombre d'enfants	nombre de couples
0	13
1	34
2	32
3	11
4	6
5	3
6	1

1. Tracer le diagramme des fréquences cumulées.
2. Calculer le nombre médian d'enfants.
3. Calculer la moyenne du nombre d'enfants.
4. S'il y a maintenant 1 couple qui a 5 enfants et 3 couples qui ont 6 enfants :

(a) Cela modifie-t-il le nombre médian d'enfants ?

(b) Cela modifie-t-il la moyenne ?

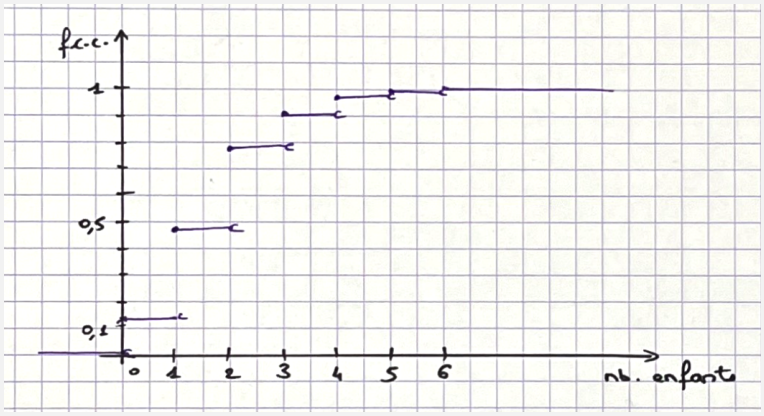
(c) Donner les nouvelles valeurs de ces grandeurs (si elles sont modifiées).

5. Déterminer la variance de cette série (non modifiée).

1. On commence par déterminer les fréquences cumulées croissantes :

nb d'enfants	nb de couples	fréquences	f.c.c.
0	13	0,13	0,13
1	34	0,34	0,47
2	32	0,32	0,79
3	11	0,11	0,90
4	6	0,06	0,96
5	3	0,03	0,99
6	1	0,01	1

On obtient le diagramme suivant.



2. L'effectif total est 100, donc la médiane est située entre la 50e et la 51e valeur.
On obtient $\boxed{M_e = 2}$.

3. On calcule la moyenne :

$$\begin{aligned}\overline{x} &= \frac{1}{100}(0\times 13 + 1\times 34 + 2\times 32 + 3\times 11 + 4\times 6 + 5\times 3 + 6\times 1)\\ &= \frac{176}{100}\\ &= 1{,}76\end{aligned}$$

Le nombre moyen d'enfants est environ 1,76

4. (a) Non, car ce changement n'affectera pas les valeurs situées en 50e et 51e positions.

(b) Oui : le nombre total des enfants a augmenté, donc la moyenne a également augmenté.

(c) On remarque que le nombre total des enfants a augmenté de 2. Donc

$$\overline{x} = \frac{178}{100} = 1{,}78$$

5. On a déjà déterminé $\overline{x} = 1{,}76$, on a donc $\overline{x}^2 = 1{,}76^2 = 2{,}8224$.

De plus, on calcule

$$\begin{aligned}\overline{x^2} &= \frac{1}{100}(0^2\times 13+1^2\times 34+2^2\times 32+3^2\times 11+4^2\times 6+5^2\times 3+6^2\times 1)\\ &= \frac{468}{100}\\ &= 4{,}68\end{aligned}$$

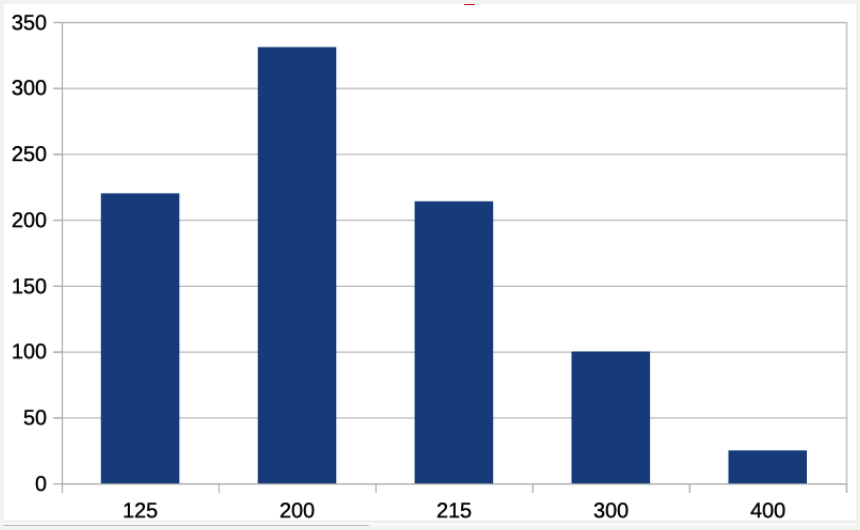
Ainsi, d'après la formule de Koenig-Huygens, la variance de cette série est

$$s_x^2 = 4{,}68 - 2{,}8224 = \boxed{1{,}8576} = 1{,}9244$$

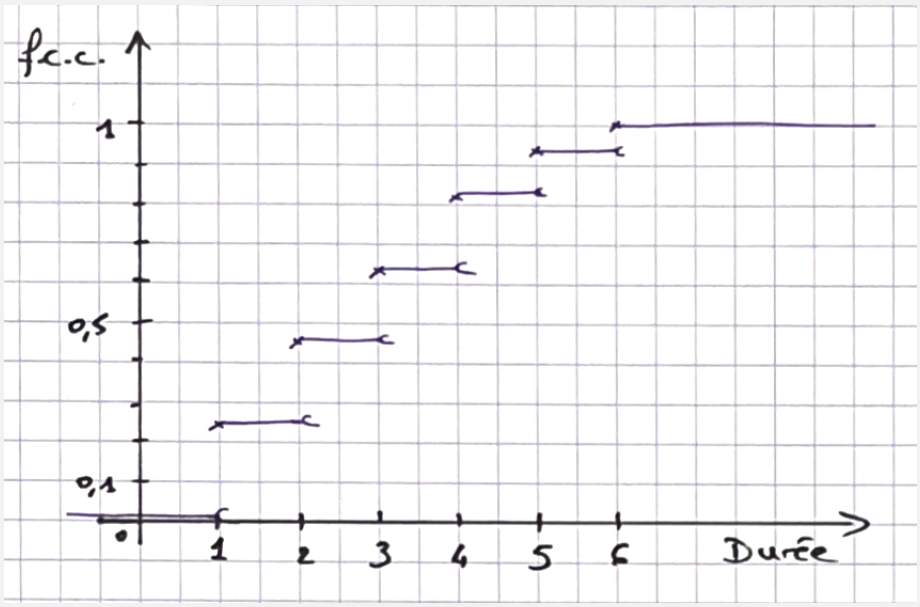
☼ **Exercice 3** (Notes) **:** On considère les notes d'étudiants à un test de langue.

Notes	100	125	200	215	300	400
Effectifs	110	220	331	214	100	25

1. Tracer le diagramme en bâtons correspondant à cette série.
2. Tracer le diagramme des fréquences cumulées.
3. Calculer la médiane de cette série.
4. Calculer Q_1 et Q_3.
5. Quelle est la note moyenne obtenue ? Quel est l'écart-type ?

1. 

Attention : l'échelle n'est pas respectée ici sur l'axe des abscisses. À faire à l'échelle pour donner une représentation plus précise des données étudiées.

2. 

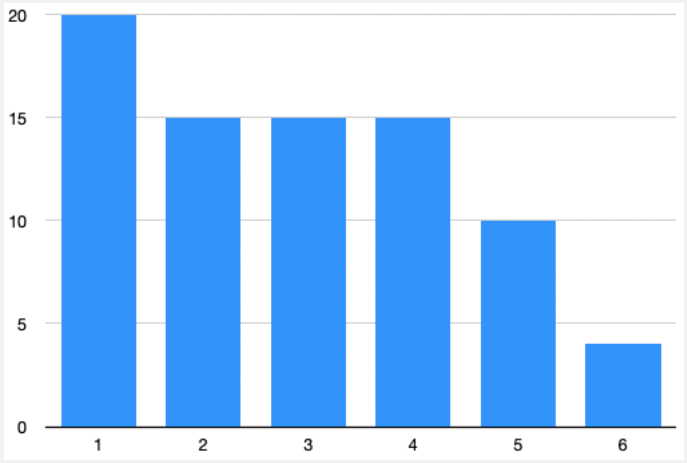
3. On trouve $\boxed{M_e = 200}$ (L'effectif total est 1000, on prend la moyenne des 500e et 501e valeurs, ou en repérant pour quelle valeur les fréquences cumulées croissantes dépassent 0,5.)
4. On trouve $\boxed{Q_1 = 125 \text{ et } Q_3 = 215}$
5. On trouve
$$\overline{x} = \frac{190\,710}{1000} = 190,71$$
On a donc $\overline{x}^2 = 36\,370,3041$. On calcule ensuite $\overline{x^2} = 40\,669,65$
On en déduit par la formule de Koenig $s_x^2 = 4\,299,3459$ puis $s_x = \sqrt{s_x^2} \approx 65,6$.
Ainsi, $\boxed{\text{la note moyenne est } 190,71 \text{ et l'écart-type est } 65,6}$.

2. En autonomie

Exercice 4 (Durée de fonctionnement d'un produit) **:** Une entreprise fabrique des produits. Elle effectue des tests sur un échantillon pour mesurer la durée de fonctionnement (arrondie en centaines d'heures). Les résultats sont donnés ci-dessous :

Durée	1	2	3	4	5	6
Effectifs	20	15	15	15	10	4

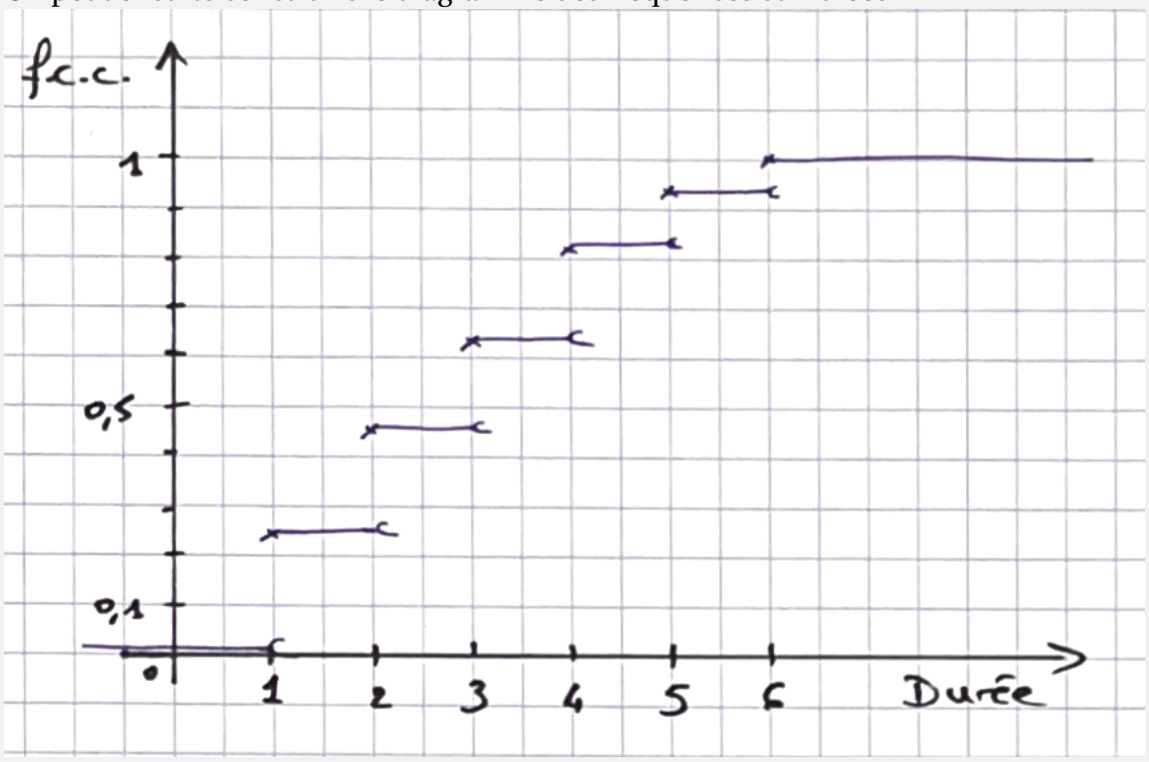
1. Tracer le diagramme en bâtons correspondant à cette série.
2. Tracer le diagramme des fréquences cumulées.
3. Déterminer la moyenne et l'écart-type de cette série.
4. Déterminer la médiane et les quartiles de cette série.
5. Dessiner la boîte à moustaches.

1. 

2. On commence par déterminer les fréquences cumulées :

Durée	1	2	3	4	5	6
Fréquences cumulées	$\frac{20}{79}$	$\frac{35}{79}$	$\frac{50}{79}$	$\frac{65}{79}$	$\frac{75}{79}$	$\frac{79}{79}=1$

On peut ensuite construire le diagramme des fréquences cumulées



3.

$$\begin{aligned}\overline{x} &= \frac{1\times 20+2\times 15+3\times 15+4\times 15+5\times 10+6\times 4}{79}\\ &= \frac{20+30+45+60+50+24}{79}\\ &= \frac{229}{79}\\ &\approx 2{,}9\end{aligned}$$

Donc la durée moyenne de fonctionnement est 290 heures environ.

Pour déterminer la variance, on commence par calculer la moyenne des carrés :

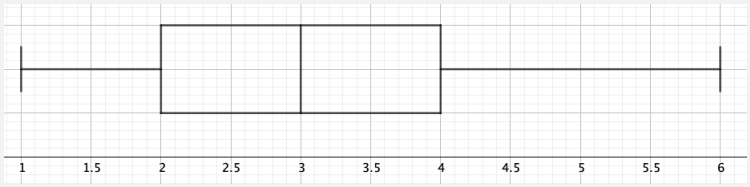
$$\begin{aligned}\overline{x^2} &= \frac{1^2\times 20+2^2\times 15+3^2\times 15+4^2\times 15+5^2\times 10+6^2\times 4}{79}\\ &= \frac{849}{79}\\ &\approx 10{,}7\end{aligned}$$

On en déduit par la formule de Koenig que

$$s_x^2 = \overline{x^2} - \overline{x}^2 = \frac{849}{79} - \left(\frac{229}{79}\right) \approx 2{,}29$$

La variance vaut donc environ 2,29.

On obtient donc comme écart-type $s_x \approx \sqrt{2{,}344} \approx 1{,}53$

4. L'effectif total est 79.
   Pour déterminer la médiane, on remarque que $79\div 2 = 39{,}5$ donc la médiane est la 40^{e} valeur.
   On a donc $M_e = 3$.
   Pour déterminer les quartiles, on calcule $\frac{1}{4}\times 79 = 19{,}75$ et $\frac{3}{4}\times 79 = 59{,}25$. Donc le premier et troisième quartiles sont respectivement les 20^{e} et 60^{e} valeurs.
   On obtient $Q_1 = 1$ et $Q_3 = 4$.
   Attention à bien prendre en compte les effectifs de chaque valeur pour déterminer médianes et quartiles.
5. 

3. Approfondissement

Exercice 5 : On considère la série statistique $(x_i, n_i)_{i\in[\![1,m]\!]}$ avec $m \geq 3$. On considère deux entiers p et q tels que $1 \leq p < q < m$. On scinde cette série en trois « sous-séries » :

— $(x_i, n_i)_{i\in[\![1,p]\!]}$ d'effectif total égal à 150, de moyenne égale à 16 et d'écart-type égal à 2.

— $(x_i, n_i)_{i\in[\![p+1,q]\!]}$ d'effectif total égal à 250, de moyenne égale à 8 et d'écart-type égal à 3.

— $(x_i, n_i)_{i\in[\![q+1,m]\!]}$ d'effectif total égal à 400, de moyenne égale à 7 et d'écart-type égal à 5.

Déterminer la moyenne et l'écart-type de la série totale.

On commence par interpréter les données des trois sous-séries :

Sous-série 1. La moyenne est 16 et l'effectif est 150, on obtient donc

$$\frac{1}{150}\sum_{i=1}^{p} n_i x_i = 16 \text{ soit } \sum_{i=1}^{p} n_i x_i = 2400$$

De même, son écart-type est 2, ce qui donne

$$2^2 = \sum_{i=1}^{p} n_i x_i^2 - 16^2 \text{ soit } \sum_{i=1}^{p} n_i x_i^2 = 39\,000$$

Sous-série 2. Avec le même raisonnement on obtient

$$\sum_{i=p+1}^{q} n_i x_i = 2000 \text{ et } \sum_{i=p+1}^{q} n_i x_i^2 = 18\,250$$

Sous-série 3. Avec le même raisonnement on obtient

$$\sum_{i=q+1}^{m} n_i x_i = 2800 \text{ et } \sum_{i=q+1}^{m} n_i x_i^2 = 29\,600$$

Calcul de $\overline{x}$ pour la série initiale. On remarque que l'effectif total est $N = 150 + 250 + 400 = 800$. Ainsi,

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{m} n_i x_i = \frac{1}{800}\left(\sum_{i=1}^{p} n_i x_i + \sum_{i=p+1}^{q} n_i x_i + \sum_{i=q+1}^{m} n_i x_i\right) = \frac{1}{800}(2400 + 2000 + 2800 = \frac{7200}{800} = 9$$

Calcul de l'écart-type σ pour la série initiale. On va appliquer la formule de Koenig-Huygens pour calculer la variance :
$\sigma^2 = \frac{1}{N}\sum_{i=1}^{m} n_i x_i^2 - (\overline{x})^2$. Ainsi,

$$\sigma^2 = \frac{1}{800}\left(\sum_{i=1}^{p} n_i x_i^2 + \sum_{i=p+1}^{q} n_i x_i^2 + \sum_{i=q+1}^{m} n_i x_i^2\right) - 9^2 = \frac{1}{800}(39\,000 + 18\,250 + 29\,600) - 81 = 27{,}5625$$

Conclusion. On obtient donc $\boxed{\overline{x} = 9 \text{ et } \sigma = \sqrt{27{,}5625} = 5{,}25}$

Exercice 6 : La répartition des notes aux tests de logique d'une école de management est donnée par le tableau suivant :

Notes	[0,4[	[4,6[	[6,8[	[8,10[	[10,12[	[12,14[	[14,16[	[16,20[
Fréquences (%)	1,08	4,81	21,03	32,45	28,00	9,50	2,88	0,24

1. Dessiner l'histogramme de cette série
2. Calculer la moyenne et la médiane de cette série. Pour effectuer ces calculs, on prendra la valeur centrale de chaque intervalle.
   L'école annonce une moyenne de 9,40. Expliquer la différence avec la valeur trouvée.

1. Ici les classes ont des largeurs différentes. Comme les aires doivent être proportionnelles, les aires des premier et dernier rectangles doivent être ajustées. Celles-ci seront à peine visibles. On peut par exemple prendre les dimensions suivantes. Ici on a choisi 1/2 cm par effectif pour toutes les classes d'amplitude 2, et 1/4 cm pour les deux classes d'amplitude 4.

Classes	$[0,4[$	$[4,6[$	$[6,8[$	$[8,10[$	$[10,12[$	$[12,14[$	$[14,16[$	$[16,20[$
Hauteurs (%)	0,27	2,40	10,51	16,22	14,00	4,750	1,44	0,06

2. **Calcul de la moyenne.** Pour calculer la moyenne, on considère le centre de chaque classe :

Valeur	2	5	7	9	11	13	15	18
Fréquences (%)	1,08	4,81	21,03	32,45	28,00	9,50	2,88	0,24

On applique la formule $\overline{x} = \sum_{i=1}^{n} f_i x_i$ pour obtenir $\boxed{\overline{x} = 9,4458}$

La différence avec la moyenne annoncée s'explique par l'approximation qui a été effectuée en regroupant les données en classes. Pour le calcul de la moyenne, on a pris les centres des classes au lieu des valeurs précises obtenues à ces tests.

Calcul de la médiane. On dresse le tableau des fréquences cumulées croissantes.

Classes	$[0,4[$	$[4,6[$	$[6,8[$	$[8,10[$	$[10,12[$	$[12,14[$	$[14,16[$	$[16,20[$
f.c.c (%)	0,0108	0,0589	0,2692	0,5937	0,8737	0,9687	0,9975	0,9999

La première classe à dépasser 0,5 est la classe $[8,10[$. Une approximation donne donc $\boxed{M_e \simeq 9}$