Correction TD 5 : Statistique Descriptive

Exercice 1.

1. $(-3-4+0+1+5+5+2-1-5+2+6+7)/12 = 15/12 = 5/4$.
2. $-5; -4; -3; -1; 0; 1; 2; 2; 5; 5; 6; 7$.
3. $n = 12$ qui est pair donc si $x_1^* \leq \cdots \leq x_{12}^*$ désigne les valeurs ordonnées, alors la médiane est $(x_6^* + x_7^*)/2 = 3/2$.
4. Si $np \in \mathbb{N}$, alors $q_{n,p} = (x_{np}^* + x_{np+1}^*)/2$. Donc $q_{12,1/4} = (x_3^* + x_4^*)/2 = -2; q_{12,1/2} = (x_6^* + x_7^*)/2 = 3/2$ et $q_{12,3/4} = (x_9^* + x_{10}^*)/2 = 5$.
5. L'étendue vaut $x_{12}^* - x_1^* = 7 - (-5) = 7 + 5 = 12$.

Exercice 2.

Partie I)

1. L'effectif est de 18 (c'est le nombre de valeurs).
2. n est pair donc on commence par ordonner les valeurs

$$4; 4; 5; 5; 5; 6; 6; 7; 7; 8; 8; 8; 9; 10; 10; 11; 13; 15.$$

   Si $x_1^* \leq \cdots \leq x_{18}^*$ désigne les valeurs ordonnées, alors la médiane est $(x_9^* + x_{10}^*)/2 = 15/2$.
3. $np = 18/4 = 9/2 \notin \mathbb{N}$ donc $q_{18,1/4} = x_{\lfloor 9/2 \rfloor + 1}^* = x_5^* = 5$.
4. L'écart interquartile vaut $q_{n,3/4} - q_{n,1/4}$. Comme $q_{18,3/4} = x_{\lfloor 27/2 \rfloor + 1}^* = x_{14}^* = 10$, on a $q_{n,3/4} - q_{n,1/4} = 10 - 5 = 5$.

Partie II)

1. À partir des fréquences cumulées, on peut déduire les fréquences relatives, qui sont respectivement : 0.15 ; 0.24-0.15 ; 0.37-0.24 ; 0.53-0.37 ; 0.79-0.53 ; 1-0.79, soit 0.15 ; 0.09 ; 0.13 ; 0.16 ; 0.26 ; 0.21. Donc le pourcentage de la valeur 25 est 16%.
2. La fréquence cumulée de 25 est 0.53 donc la médiane est 25.
3. La fréquence cumulée de 23 est 0.24 et celle de 24 est 0.37, donc le premier quartile est 24.

Partie III)

1. L'étendue vaut $20 - 4 = 16$.
2. La médiane vaut 15.
3. Le 1er quartile vaut 10.
4. L'écart interquartile vaut $16 - 10 = 6$.
5. La moitié des valeurs sont plus petites que la médiane, i.e., que 15, mais pas que 10.
6. 75% des valeurs sont inférieures au 3ème quartile, donc à 16.
7. 75%, resp. 25%, des valeurs sont inférieures au 3ème quartile, resp. au 1er quartile, donc à 16, resp. à 10. Donc la moitié des valeurs sont comprises entre 10 et 16.

Exercice 3.

Remarque : Chaque affirmation est composée de deux assertions. Il suffit que l'une soit fausse pour que l'affirmation le soit.

Si x_i désigne les notes avant transformation et y_i après transformation, on a clairement

- $\overline{y}_n = \frac{1}{n}\sum_{i=1}^n y_i = \frac{11}{10}\overline{x}_n$;
- $s_y^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y}_n)^2 = \frac{121}{100}s_x^2$ et donc $s_y = \frac{11}{10}s_x$;
- médiane des y $= \frac{11}{10}\times$ médiane des x.
- quartiles de $y = \frac{11}{10}\times$ quartiles de x, donc il en va de même de l'écart interquartile.

1. Faux, la moyenne empirique sera multipliée par 11/10.
2. Faux, la moyenne empirique sera multipliée par 11/10.
3. Faux, l'écart-type empirique sera multipliée par 11/10.
4. Faux, la variance empirique augmentera de 21%.
5. Vrai.
6. Faux, la moyenne empirique sera multipliée par 11/10.
7. Faux, la médiane sera multipliée par 11/10.
8. Faux, la variance empirique augmentera de 21%.
9. Faux, l'écart interquartile sera multiplié par $\frac{11}{10}$.

Exercice 4.

1. Aucune des réponses précédentes. En effet, si le nouvel échantillon est noté : $y_1 = x_1, \ldots, y_n = x_n, y_{n+1} = M_x$ où M_x désigne la médiane de l'échantillon $x_1, ..., x_n$, alors

$$\begin{aligned}
\overline{y}_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1} y_i \\
&= \frac{1}{n+1}\left(\sum_{i=1}^{n} x_i + M_x\right) \\
&= \frac{1}{n+1}\left(n\,\frac{1}{n}\sum_{i=1}^{n} x_i + M_x\right) \\
&= \frac{1}{n+1}\left(n\,\overline{x}_n + M_x\right) \\
&= \frac{n}{n+1}\overline{x}_n + \frac{M_x}{n+1}.
\end{aligned}$$

   Donc $\overline{y}_{n+1} \geq \overline{x}_n$ ssi $M_x \geq \overline{x}_n$.

2. Faux, la variance empirique $s_x^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}_n)^2$, donc on ne peut rien dire.
3. Ne change pas, car si x_i représente le résultat du test pour $i = 1, \ldots, 10$, avec $\overline{x}_{10} = 75$ et si maintenant on note $y_i = x_i - 75$. Alors la variance empirique basée sur les y_i est la même que celle basée sur les x_i car $\overline{y}_n = \overline{x}_n - 75$ et donc

$$s_y^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y}_n)^2 = \frac{1}{n}\sum_{i=1}^n [x_i - \overline{x}_n)^2 = s_x^2.$$

4. $y_i = x_i/100$ donc $\overline{y}_n = \overline{x}_n/100$.
5. $y_i = x_i/100$ donc $s_y^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y}_n)^2 = \frac{1}{n}\sum_{i=1}^n \left(\frac{x_i}{100} - \frac{\overline{x}_n}{100}\right)^2 = \frac{s_x^2}{10\,000}$.