

**Élément de statistique : évaluation formative**

**Partie 1 (C1/C2/C3/C4)**

**Nom :**

**Prénom :**

**Consignes :**

- Toutes vos réponses doivent être **justifiées** et votre argumentation doit convaincre le lecteur. L'argumentation doit être courte (une phrase ou deux maximum).
- Votre écriture doit être le plus lisible possible.
- Les réponses finales doivent figurer dans les encadrés prévus à cet effet. Vous pouvez utiliser tous les versos de toutes les pages comme brouillon.

**Exercice 1**

Voici un tableau qui informe du nombre de places de concert (variable  $X_i$ ) achetées au cours de l'année 2018 par un échantillon représentatif de 10 étudiants belges du secondaire, ainsi que leur âge (variable  $Y_i$ ).

<b>i</b>	<b>(<math>X_i</math>)</b>	<b>(<math>Y_i</math>)</b>
1	3	16
2	2	15
3	4	18
4	5	21
5	2	16
6	0	14
7	1	15
8	2	15
9	2	16
10	3	17

**1.1 Quel est le type de la variable  $X_i$  ? Soyez aussi précis(e) que possible, et justifiez votre réponse.**

Quantitatif discret. Réponses chiffrées (et sens intrinsèque des chiffres), et pas de décimale possible (chiffres « ronds » obligatoires).

**1.2 Quel est le type de la variable  $Y_i$  ? Soyez aussi précis(e) que possible, et justifiez votre réponse.**

Quantitatif continu. Réponses chiffrées (et sens intrinsèque des chiffres), et décimales possibles.

**1.3 Réalisez un tableau de fréquences complet de la variable  $X_i$ . N'oubliez pas de nommer les colonnes en utilisant les notations mathématiques adéquates.**

Attention : dans chaque colonne, j'ai déjà placé une valeur, afin de vous amener à savoir quelle information doit se trouver dans quelle colonne du tableau. Respectez bien l'ordre d'apparition des colonnes que je vous impose.

<b>J</b>	<b><math>x_j</math></b>	<b><math>n_j</math></b>	<b><math>f_j</math></b>	<b><math>N_j</math></b>	<b><math>F_j</math></b>
1	0	1	10%	1	10%
2	1	1	10%	2	20%
3	2	4	40%	6	60%
4	3	2	20%	8	80%
5	4	1	10%	9	90%
6	5	1	10%	10	100%

**1.4 Calculez la moyenne, le mode, la médiane et l'écart moyen absolu de la variable  $X_i$ , en utilisant la formule de votre choix. Dans l'espace disponible à la page 3, dites brièvement la méthode/formule que vous avez choisie et détaillez vos calculs. Retransmettez ensuite la réponse finale dans le tableau ci-dessous. N'arrondissez pas vos réponses.**

Attention. Le calcul est tout aussi important que la réponse finale : juste fournir la réponse sans montrer le calcul pour y arriver ne sera pas accepté ! Pour chaque indicateur, expliquez en quelques mots la formule/méthode choisie, et détaillez vos calculs de manière aussi structurée que possible.

	<b>Variable <math>X_i</math></b>
<b>Moyenne</b>	<b>2,4</b>
<b>Mode</b>	<b>2</b>
<b>Médiane</b>	<b>2</b>
<b>EMA</b>	<b>1,08</b>

### Espace de calcul

Moyenne : je propose trois solution de réponse, mais une seule suffit (vous choisissez celle avec laquelle vous êtes le(la) plus à l'aise).

**Au départ des données brutes :**  $\frac{3+2+4+5+2+0+1+2+2+3}{10} = \frac{24}{10} = 2,4$

**Au départ des fréquences absolues :**  $\frac{1 \times 0 + 1 \times 1 + 4 \times 2 + 2 \times 3 + 1 \times 4 + 1 \times 5}{10} = \frac{24}{10} = 2,4$

**Au départ des fréquences relatives :**  $0,1 \times 0 + 0,1 \times 1 + 0,4 \times 2 + 0,2 \times 3 + 0,1 \times 4 + 0,1 \times 5 = 2,4$

Mode : 2 (car il s'agit de la valeur associée à la fréquence absolue/relative la plus élevée)

Médiane : je propose deux solutions de réponse, mais une seule suffit (vous choisissez celle avec laquelle vous êtes le(la) plus à l'aise).

**Au départ des données brutes :**

1) Trier les observations par ordre croissant et leur associer un rang :

$X_i$	0	1	2	2	2	2	3	3	4	5
Rang	1	2	3	4	5	6	7	8	9	10

2) Calculer le rang médian  $= \frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5,5$  (soit entre les rangs 5 et 6)

3) Au départ du tableau de fréquences :  $\frac{2+2}{2} = 2$

**Au départ des fréquences relatives :** regarder la première modalité de  $x_j$  dont la fréquence relative cumulée dépasse 50%  $\rightarrow 2$

**ATTENTION** : bien que ça ne soit pas le cas dans l'exemple ici, n'oubliez pas que si la fréquence relative cumulée d'une modalité vaut EXACTEMENT 50%, la valeur de la médiane sera la moyenne entre cette modalité et la modalité qui vient juste après (comme expliqué dans le cours).

EMA : même remarque : une seule méthode suffit

**Au départ des données brutes :**  $\frac{\sum_{i=1}^n |X_i - \bar{X}|}{10} = \frac{10,8}{10} = 1,08$

i	( $X_i$ )	( $X_i - \bar{X}$ )	$X_i - \bar{X}$
1	3	$3 - 2,4 = 0,6$	0,6
2	2	$2 - 2,4 = -0,4$	0,4
3	4	$4 - 2,4 = 1,6$	1,6
4	5	$5 - 2,4 = 2,6$	2,6
5	2	$2 - 2,4 = -0,4$	0,4
6	0	$0 - 2,4 = -2,4$	2,4
7	1	$1 - 2,4 = -1,4$	1,4
8	2	$2 - 2,4 = -0,4$	0,4
9	2	$2 - 2,4 = -0,4$	0,4
10	3	$3 - 2,4 = 0,6$	0,6

**Au départ des fréquences absolues :**  $\frac{\sum_{j=1}^k n_j |X_j - \bar{X}|}{10} = \frac{1 \times 2,4 + 1 \times 1,4 + 4 \times 0,4 + 2 \times 0,6 + 1 \times 1,6 + 1 \times 2,6}{10} = \frac{10,8}{10} = 1,08$

**Au départ des fréquences relatives :**  $\sum_{j=1}^k f_j |X_j - \bar{X}| = 0,1 \times 2,4 + 0,1 \times 1,4 + 0,4 \times 0,4 + 0,2 \times 0,6 + 0,1 \times 1,6 + 0,1 \times 2,6 = 1,08$

**1.5 Pour chaque variable, calculez la variance et l'écart-type, en utilisant la formule de votre choix. Utilisez l'espace disponible ci-dessous pour détailler vos calculs, et ne transmettez que la réponse finale dans le tableau du bas de la page. Arrondissez vos réponses finales à 2 décimales. Attention. Même remarque qu'au point précédent.**

### Espace de calcul

#### Variance de $X_i$ :

Au départ des données brutes :  $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{10} = \frac{18,4}{10} = 1,84$

i	( $X_i$ )	( $X_i - \bar{X}$ )	( $X_i - \bar{X}$ ) <sup>2</sup>
1	3	0,6	0,36
2	2	-0,4	0,16
3	4	1,6	2,56
4	5	2,6	6,76
5	2	-0,4	0,16
6	0	-2,4	5,76
7	1	-1,4	1,96
8	2	-0,4	0,16
9	2	-0,4	0,16
10	3	0,6	0,36

Au départ des fréquences absolues :  $\frac{\sum_{j=1}^k n_j (x_j - \bar{X})^2}{10} = \frac{1 \times 5,76 + 1 \times 1,96 + 4 \times 0,16 + 2 \times 0,36 + 1 \times 2,56 + 1 \times 6,76}{10} = \frac{18,4}{10} = 1,84$

Au départ des fréquences relatives :  $\sum_{j=1}^k f_j (x_j - \bar{X})^2 = 0,1 \times 5,76 + 0,1 \times 1,96 + 0,4 \times 0,16 + 0,2 \times 0,36 + 0,1 \times 2,56 + 0,1 \times 6,76 = 1,84$

Écart type de  $X_i$  :  $\sqrt{1,84} = 1,36$

#### Variance de $Y_i$ :

Au départ des données brutes :  $\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{10} = \frac{36,1}{10} = 3,61$

i	( $Y_i$ )	( $Y_i - \bar{Y}$ )	( $Y_i - \bar{Y}$ ) <sup>2</sup>
1	16	-0,3	0,09
2	15	-1,3	1,69
3	18	1,7	2,89
4	21	4,7	22,09
5	16	-0,3	0,09
6	14	-2,3	5,29
7	15	-1,3	1,69
8	15	-1,3	1,69
9	16	-0,3	0,09
10	17	0,7	0,49

Au départ des fréquences absolues :  $\frac{\sum_{j=1}^k n_j (y_j - \bar{Y})^2}{10} = \frac{1 \times 5,29 + 3 \times 1,69 + 3 \times 0,09 + 1 \times 0,49 + 1 \times 2,89 + 1 \times 22,09}{10} = \frac{36,1}{10} = 3,61$

Au départ des fréquences relatives :  $\sum_{j=1}^k f_j (y_j - \bar{Y})^2 = 0,1 \times 5,29 + 0,3 \times 1,69 + 0,3 \times 0,09 + 0,1 \times 0,49 + 0,1 \times 2,89 + 0,1 \times 22,09$

Écart type de  $X_i$  :  $\sqrt{3,61} = 1,9$

	Variable $X_i$	Variable $Y_i$
<b>Variance</b>	<b>1,84</b>	<b>3,61</b>
<b>Écart-type</b>	<b>1,36</b>	<b>1,90</b>

**1.6 Déterminer la valeur du coefficient de corrélation  $r$  de Pearson pour déterminer la relation entre les variables  $X_i$  et  $Y_i$ . Arrondissez votre réponse à 3 décimales.**

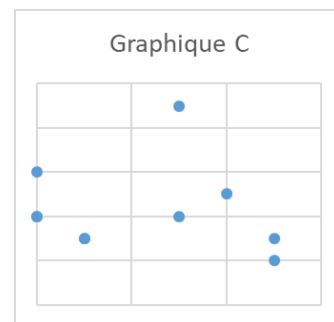
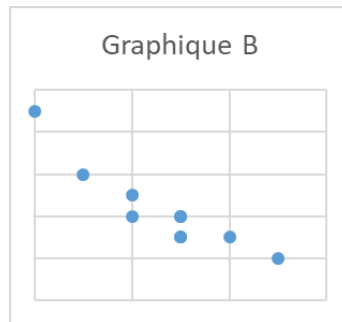
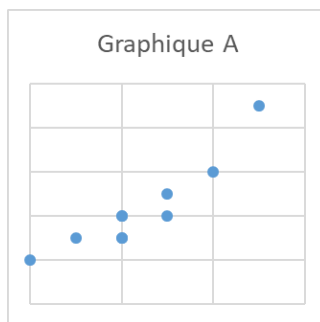
Covariance = 2,38

Ecart-type 1 = 1,36

Ecart-type 2 = 1,90

Corrélation  $r$  de Pearson  $\approx 0,923$  (en fonction de vos arrondis, vous n'aurez peut-être pas exactement la même réponse !)

**1.7 Parmi les trois graphiques ci-dessous, lequel représente la relation entre  $X_i$  et  $Y_i$  ? Justifiez.**



Réponse : graphique A (puisque c'est celui qui montre une forte relation positive entre les variables)

**1.8 Comment interprétez la relation entre les variables  $X_i$  et  $Y_i$  ? Indiquez une croix devant la réponse correcte.**

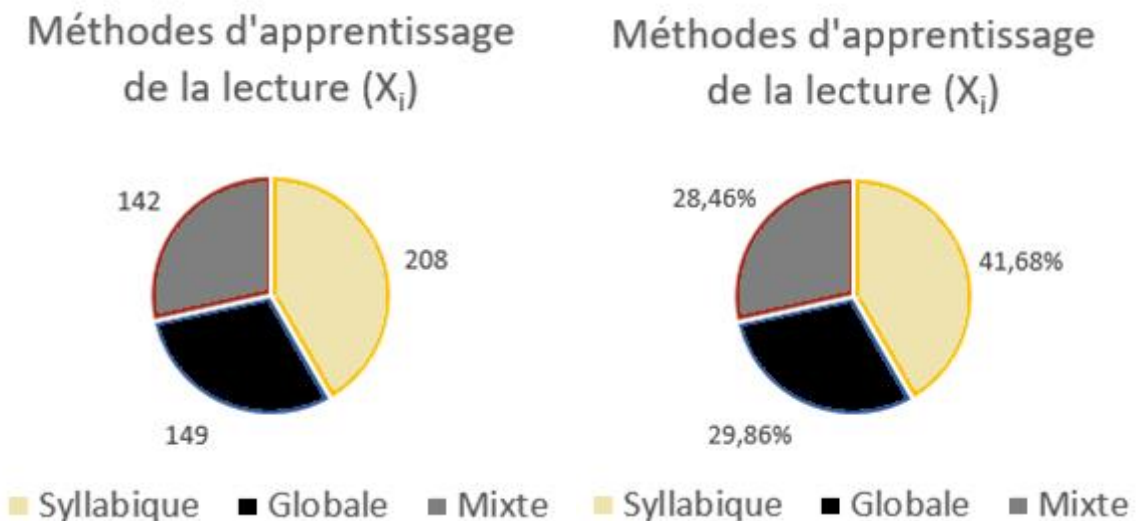
	Corrélation positive moyenne
	Corrélation négative moyenne
X	Forte corrélation positive
	Forte corrélation négative

## Exercice 2

Il existe plusieurs méthodes pour apprendre aux enfants à lire. Parmi ces méthodes il y a :

- La méthode syllabique (on part du détail vers l'ensemble : les enfants doivent détecter des syllabes pour arriver à former des mots)
- La méthode globale (on part de l'ensemble vers le détail : les enfants doivent connaître un mot en entier, avant d'en analyser les éléments qui le constituent)
- La méthode mixte (un mélange des méthodes syllabique et globale)

On a demandé à un ensemble d'instituteurs, provenant de diverses écoles un peu partout en Europe, la méthode qu'ils utilisent en classe avec leurs élèves. Le graphique ci-dessous représente les résultats obtenus :



**2.1 Combien vaut n ? Montrez votre calcul.**

$$142+149+208 = 499$$

**2.2 Quel est le type de la variable représentée ? Soyez aussi précis(e) que possible, et justifiez votre réponse.**

Variable qualitative nominale. Les individus sont placés dans des catégories, et pas d'ordre logique entre les catégories.

**2.3 Combien la variable étudiée a-t-elle de modalités ? Donnez-en le nombre, et citez-les.**

3 modalités : méthode syllabique, méthode globale et méthode mixte

**2.4 Comment appelle-t-on la représentation graphique ci-dessus ? Expliquez brièvement un inconvénient de ce type de graphique, et citez un autre type de graphique qui permet de pallier cet inconvénient.**

Graphique réalisé :

Diagramme circulaire.

Inconvénient : Difficile de distinguer deux catégories lorsqu'elles sont associées à des fréquences relativement similaires.

Alternative : le diagramme en bâtons.

**2.5 Quel est le type de méthode le plus utilisé par les instituteurs de l'échantillon ?**

Méthode syllabique.

**2.6 Comment appelle-t-on l'indicateur que vous avez dû déterminer au point 2.5 ?**

Le mode de la série

**2.7 Quels sont les deux autres indicateurs de tendance centrale vus au cours ? Citez-les.**

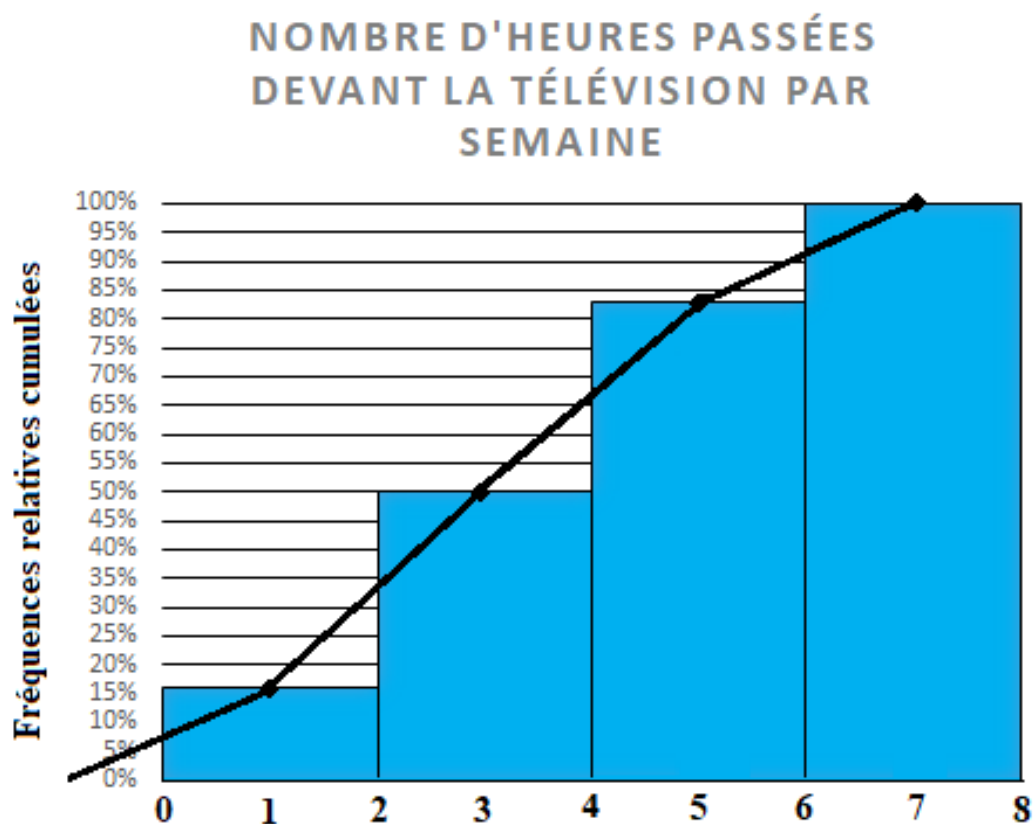
Moyenne et médiane

**2.8 Peut-on calculer, pour la variable étudiée, les deux indicateurs cités au point 2.7 ? Justifiez votre réponse.**

Non, car il s'agit d'une variable nominale !

### Exercice 3

Voici un tableau de fréquences et une représentation graphique qui représentent les mêmes données.



**Tableau de fréquences du nombre d'heures passées devant la télévision par semaine**

j	$x_j$	Centre de classe	$n_j$	$f_j$	$N_j$	$F_j$
1	0 – 2	1	32	32	16%	16%
2	2 – 4	3	68	100	34%	50%
3	4 – 6	5	66	166	33%	83%
4	6 – 8	7	34	200	17%	100%

**3.1 Le tableau de fréquence n'est pas complet : il manque le nom de 4 colonnes. Complétez-le, en utilisant les notations mathématiques adéquates.**



**3.2 Approximez les valeurs du premier et troisième quartile, grâce à la technique de l'interpolation linéaire.**

Valeur de Q1 : 1,53

2

Espace calcul :

$$50\% - 16\% = 34\% \rightarrow 3 - 1 = 2$$

$$25\% - 16\% = 9\% \rightarrow \frac{2}{34} \times 9 = 0,53 \rightarrow Q1 = 1 + 0,53 = \underline{1,53}$$

Valeur de Q3 : 4,52

Espace calcul :

$$83\% - 50\% = 33\% \rightarrow 3 - 1 = 2$$

$$75\% - 50\% = 25\% \rightarrow \frac{2}{33} \times 25 = 1,52 \rightarrow Q3 = 3 + 1,52 = \underline{4,52}$$

**3.3 Approximez la valeur de l'écart-interquartile.**

$$4,52 - 1,53 = 2,99$$

**3.4 Quel est le pourcentage d'individus de la base de données étudiée dont le temps passé devant la télévision se situe entre les valeurs des premier et le troisième quartile ? Indiquez une croix devant la réponse correcte.**

	25%
<b>X</b>	50%
	75%

**3.5 Quel est le pourcentage d'individus de la base de données étudiée dont le temps passé devant la télévision est inférieur à la valeur du premier quartile ? Indiquez une croix devant la réponse correcte.**

<b>X</b>	25%
	50%
	75%