

Chers élèves,

Tout d'abords, j'espère que vous vous portez bien ainsi que vos proches. Cette période est loin d'être simple à gérer, tant au niveau familial qu'au niveau scolaire.

Afin de ne pas perdre vos bonnes habitudes mathématiques, nous vous avons préparé un nouveau dossier, le dernier dossier relatif aux statistiques à une variable.

Lisez le attentivement, tentez de répondre aux questions, et nous le corrigerons ensemble lors du retour en classe!

Si vous êtes bloqué, n'hésitez pas à me contacter par mail, je tenterais de vous aiguiller au mieux!

En attendant, soyez prudents, prenez soin de vous et de vos proches,

Mme Dewandre

Dewandre.prof@gmail.com

DOSSIER 5 - Indices de dispersion

1 Savoirs

1.1 Définitions à partir de variables discrètes

Exemple

On a relevé une série de 30 températures. En voici le tableau recensé et le diagramme des effectifs cumulés.

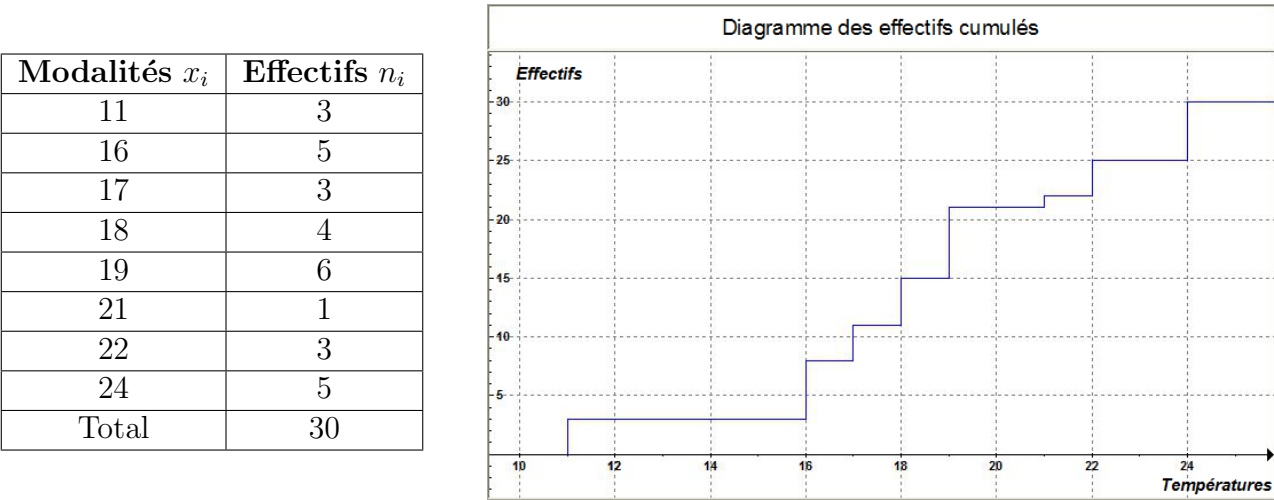


FIGURE 1 – Tableau recensé et diagramme des effectifs cumulés d’une série de température

Les paramètres de dispersion permettent de rendre compte de la manière dont les données s’étendent autour de la valeur moyenne.

- L’**étendue** d’une série statistique est la différence entre la plus grande et la plus petite valeur de la série.

Dans notre exemple, elle vaut $24 - 11 = 13\text{ }^{\circ}\text{C}$.

- L’**intervalle interquartile** d’une série statistique est l’intervalle $[Q_1; Q_3]$. Il contient la moitié de la population (autour de la médiane). Il n’est pas strictement un indice de dispersion puisqu’il renseigne la position des 1^{er} et 3^e quartiles.

L’**écart interquartile** est la largeur de l’intervalle interquartile. Il vaut $Q_3 - Q_1$.

Dans notre exemple , cela donne $21 - 16 = 5\text{ }^{\circ}\text{C}$.

- Pour obtenir une valeur qui rende compte de l'écart de toutes les données par rapport à la moyenne, on utilise la variance. La **variance** V d'une série statistique est la moyenne des carrés des écarts à la moyenne de toutes les valeurs de la série.

$$V = \frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^2 \text{ où } p \text{ est le nombre de modalités de la série}$$

Dans notre exemple, la moyenne est de 18,57, la variance sera donc calculée comme suit :

$$V = \frac{1}{30} \cdot ((3 \cdot (11 - 18,57)^2) + (3 \cdot (17 - 18,57)^2) + \dots + (3 \cdot (22 - 18,57)^2) + (5 \cdot (24 - 18,57)^2))$$

$$V = 26,61 \text{ } (^{\circ}C)^2$$

Cependant, la variance n'a pas les mêmes dimensions que les données puisqu'elle utilise leur carré. **Pour retrouver un paramètre qui a plus de sens**, on prend la racine carrée positive de la variance.

- L'**écart-type** σ est la racine carrée positive de la variance.

$$\sigma = \sqrt{V} = \sqrt{\frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^2}$$

Dans notre exemple, on trouve $\sigma = 5,16 \text{ } ^{\circ}C$.

Pour faire le calcul autrement qu'en utilisant le menu « stat » d'une calculatrice ou un tableur, on va rajouter trois colonnes à notre tableau de calculs. La première donne les écarts à la moyenne $(x_i - \bar{x})$, la seconde le carré des écarts à la moyenne et la troisième le produit des carrés des écarts par l'effectif $(n_i \cdot (x_i - \bar{x})^2)$.

On fait la somme (ou le total) des valeurs trouvées dans la dernière colonne. En la divisant par l'effectif total, on trouve la variance.

Complète le tableau et vérifie que tu obtiens la bonne valeur de la variance.

Modalités	Effectifs	Fréquences	Eff. cumulées	Fréq. cumulées	Écarts à la moyenne	Carrés des écarts	
x_k	n_k	f_k	N_k	F_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	$n_k \cdot (x_k - \bar{x})^2$
11	3	0,1	3	0,1	$11 - 18,57 = -7,57$	$(-7,57)^2 = 57,3$	$3 \times 57,3 = 171,9$
16	5	0,167	8	0,267			
17	3	0,1	11	0,367			
18	4	0,133	15	0,5			
19	6	0,2	21	0,7			
21	1	0,033	22	0,733			
22	3	0,1	25	0,833			
24	5	0,167	30	1			
Total	30	/	/	/	/	/	

1.2 Application à une variable continue

L'étendue d'une série statistique se calcule, en principe, sur base des données brutes. Si on ne possède que le tableau groupé en classe on utilisera les valeurs extrêmes des classes extrêmes. Pour calculer l'écart-type d'une série dont les modalités sont regroupées en classes, on utilise (comme pour le calcul de la moyenne) les centres c_i des classes à la place des x_i .

Exercice guidé

Le directeur d'un supermarché a réalisé une enquête sur le temps d'attente aux caisses dont voici les résultats.

Temps d'attente (min)	[0 ;5[	[5 ;10[	[10 ;15[	[15 ;20[	[20 ;25[	[25 ;30]
Effectifs	18	27	45	30	22	8

En utilisant le tableau de calculs (à compléter) et le polygone des fréquences cumulées (à réaliser), détermine l'étendue, l'intervalle interquartile, l'écart interquartile et l'écart-type de la série. N'oublie pas qu'il te faut calculer la moyenne pour compléter les dernières colonnes du tableau. Vérifie qu'elle vaut $13,67 \text{ min}$.

Tu dois trouver $V = 46,8 \text{ min}^2$ et $\sigma = 6,84 \text{ min}$

Classes modales	Centres de classes	Effectifs	Fréquences	Eff. cumulées	Fréq. cumulées	Écarts à la moyenne	Carrés des écarts	
x_k	c_k	n_k	f_k	N_k	F_k	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	$n_k \cdot (x_k - \bar{x})^2$
$[0; 5[$	2,5	18	0,12	18	0,12	$2,5 - 13,67 = -11,17$	$(-11,17)^2 = 124,77$	2245,84
$[5; 10[$	7,5	27	0,18	45	0,3			
$[10; 15[$	12,5	45	0,3	90	0,6			
$[15; 20[$		30						
$[20; 25[$		22						
$[25; 30[$		8						
Total		150	/	/	/	/	/	

2 Applications

- La veille du début du confinement, un magasin d'électroménagers a classé ses clients d'après le montant de leurs achats. Le tableau de résultats est donné ci-dessous.

Classes	$[0; 50[$	$[50; 100[$	$[100; 150[$	$[150; 200[$	$[200; 250[$	$[250; 300[$	$[300; 350[$	$[350; 400[$	$[400; 450[$	$[450; 500[$
Effectifs	5	10	20	25	30	35	75	130	100	70

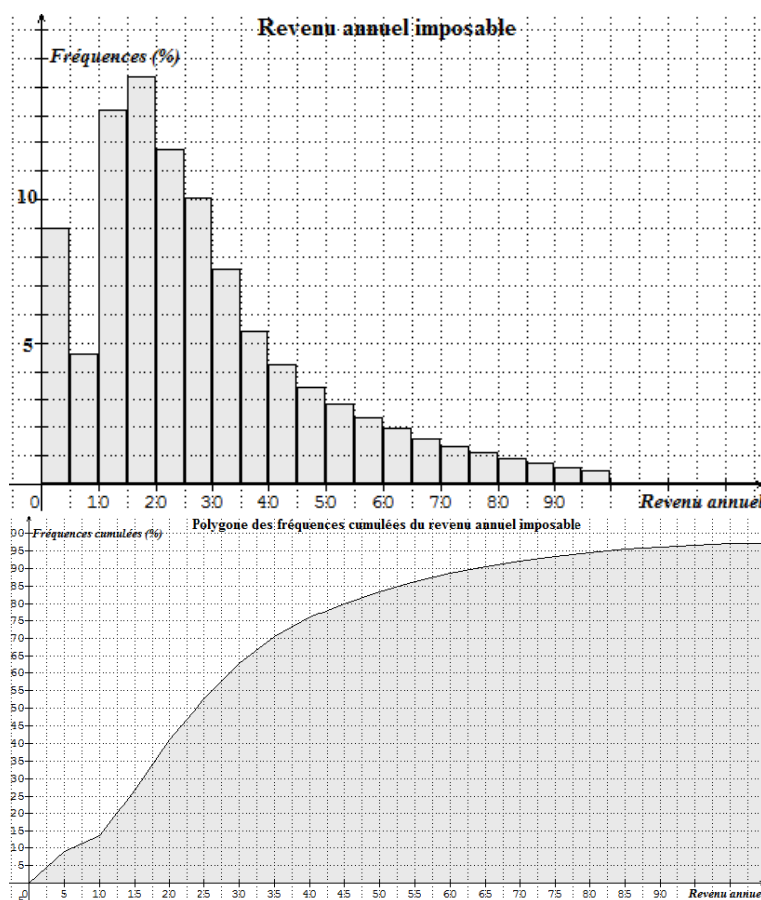
- Compare le mode, la moyenne et la médiane. Explique les différences entre ces trois valeurs de position.
- Détermine la valeur des trois indices de dispersions suivants : étendue, écart interquartile et écart-type.
- Réalise un histogramme avec ces données et places-y les éléments calculés aux deux premières questions. Attention, les indices de dispersion ne doivent pas être représentés comme les valeurs de position.

2. Afin de déterminer comment mettre en place une nouvelle taxe permettant de financer le milieu médical, on souhaite étudier les revenus des belges, une source de données précises vient des déclarations fiscales. On peut alors examiner le revenu annuel net imposable.

Le revenu annuel net imposable a ici été regroupé en classe de 5000 € de large. La première classe reprend les revenus annuel inférieurs à 5000 € et donc également tous les cas où il n'y a pas de revenu imposable. La dernière classe reprend tous les revenus supérieurs à 100 milliers d'euros.

Attention, pour la dernière classe, Il a été choisi de mettre le maximum de cette 300 (même si en réalité certaines déclarations dépassent sans soucis ce montant).

En utilisant les graphes ci-dessous, réponds aux différentes questions.



- Quel est l'intervalle interquartile ? Explique ce que cela signifie concrètement dans ce cas-ci.
- On a calculé la variance de la série qui vaut 1187,83. Déduis-en l'écart-type sans oublier de donner son unité.
- Que vaut l'écart interquartile ? Compare-le avec l'écart-type.
- Calcule **approximativement** le revenu moyen. Compare-le avec le mode ainsi qu'avec le revenu médian.

3 Exercices récapitulatifs

1. Afin de déterminer la charge de travail d'une pharmacie de garde, une étude a enregistré la fréquentation de 100 pharmacies de garde en Wallonie pendant les 10 dernières nuits entre 20h et 6h du matin. Les résultats sont les suivants :

Nombre d'appels	0	1	2	3	4	5	6	7	8	9	10	11
Nombre de nuits	14	70	155	185	205	150	115	65	30	5	1	5

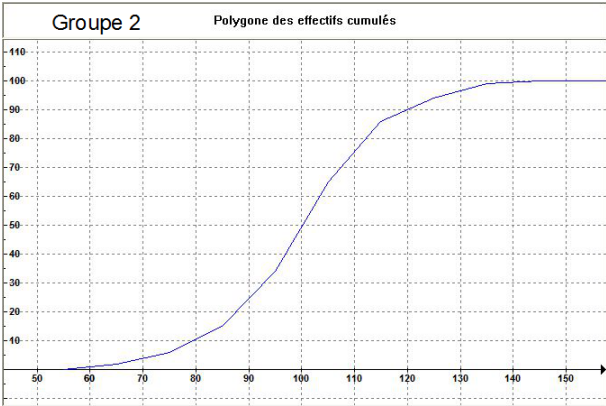
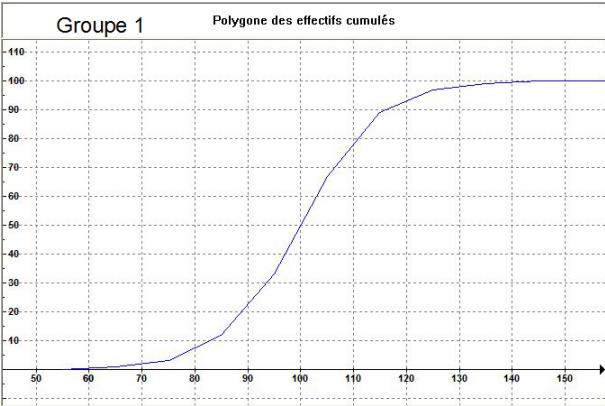
- En moyenne combien de personnes se déplacent en pharmacie de garde par nuit ?
 - Que vaut l'écart interquartile et l'écart type de cette série ?
 - Un jeune pharmacien affirme à ces amis que dans 75% des gardes il y a au moins 5 clients sur la nuit. Que penses-tu de son opinion ?
2. On a demandé à un groupe de 220 élèves de 4^e combien d'heures ils ont regardé la télévision chaque semaine. Leurs réponses ont été consignées dans le tableau ci-dessous. À l'aide de cette information, calcule la moyenne et l'écart-type des heures pendant lesquelles les 220 élèves ont regardé la télévision. Dessine également un histogramme et un polygone cumulatif des fréquences. Déduis-en le mode et la médiane de la série statistique.

Heures	Nombre d'élèves
10 à 14	2
15 à 19	12
20 à 24	23
25 à 29	60
30 à 34	77
35 à 39	38
40 à 44	8

Question Bonus : En période de confinement est ce que ces valeurs évoluent ?
Effectue une étude statistiques sur base de ton expérience et de celles de 4 autres amis/famille

3. Deux groupes de personnes ont passé des tests de Q.I.

QI	Groupe 1	Groupe 2
[55,65[	1	2
[65,75[	2	4
[75,85[	9	9
[85,95[	21	19
[95,105[	34	31
[105,115[	22	21
[115,125[	8	8
[125,135[	2	5
[135,145]	1	1



Pour chaque groupe :

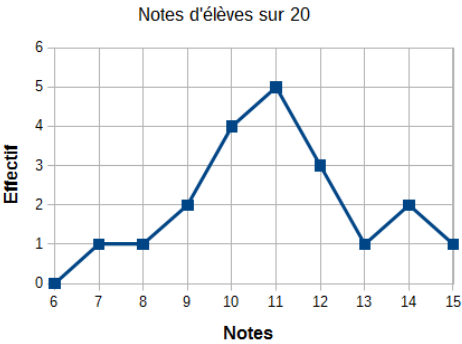
- a) Détermine et donne la signification concrète du mode, de la médiane, des premiers et troisième quartile ainsi que de l'intervalle interquartile.
- b) Combien de quotients sont inférieurs à 115 ?
- c) Détermine la moyenne et l'écart-type sachant que la variance vaut 179 pour le premier groupe et 237 pour le second. Que signifient ces chiffres ?

Compare les deux groupes. Lequel est le plus homogène ? Justifie.

4. A propos de devoirs

On considère le diagramme donné ci-contre qui reprend les résultats (sur 20) obtenus à un devoir de statistiques. Les affirmation suivantes sont-elles vraies ou fausses ? Jutifier.

- a) La classe est surchargée.
- b) 40% des élèves ont une note inférieure ou égale à 10.
- c) La moyenne de la classe est égale à 11.
- d) Il y a autant de notes qui sont supérieures à la moyenne que de notes qui lui sont inférieures.



5. Usure de pneus

On a calculé le nombre de milliers de km parcourus par 10 pneus des chacune des marques A et B avant usure. Les résultats ci-dessous ont été obtenus.

A	25	28	26	34	30	24	28	22	27	23
B	31	29	24	28	20	34	27	30	23	22

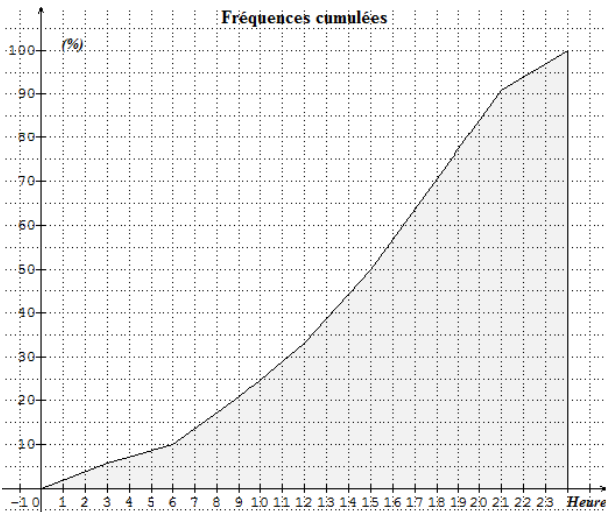
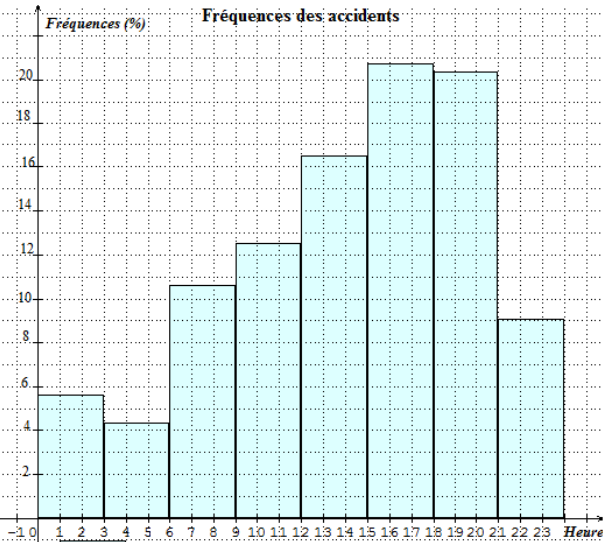
- a) Compare la qualité de pneus des deux marques en utilisant la moyenne, la médiane et les quartiles. Quelle marque conseillerais-tu ? Argumente.
- b) On a calculé l'écart-type des deux séries : 3,37 pour la marque A et 4,21 pour la marque B. Ces informations changent-elles ton choix de marque ? Pourquoi ?

4 Exercices bonus

1. Accidents

On a relevé le nombre d'accidents de la route pour différentes heures de la journée. On souhaite dégager les tendances essentielles liées à ces informations grâce aux données et aux deux graphiques qui en découlent.

Tranche horaire	[0; 3[	[3; 6[	[6; 9[	[9; 12[	[12; 15[	[15; 18[	[18; 21[	[21; 24[
Nombre d'accidents	8155	6258	15284	18006	23703	29759	29172	13022

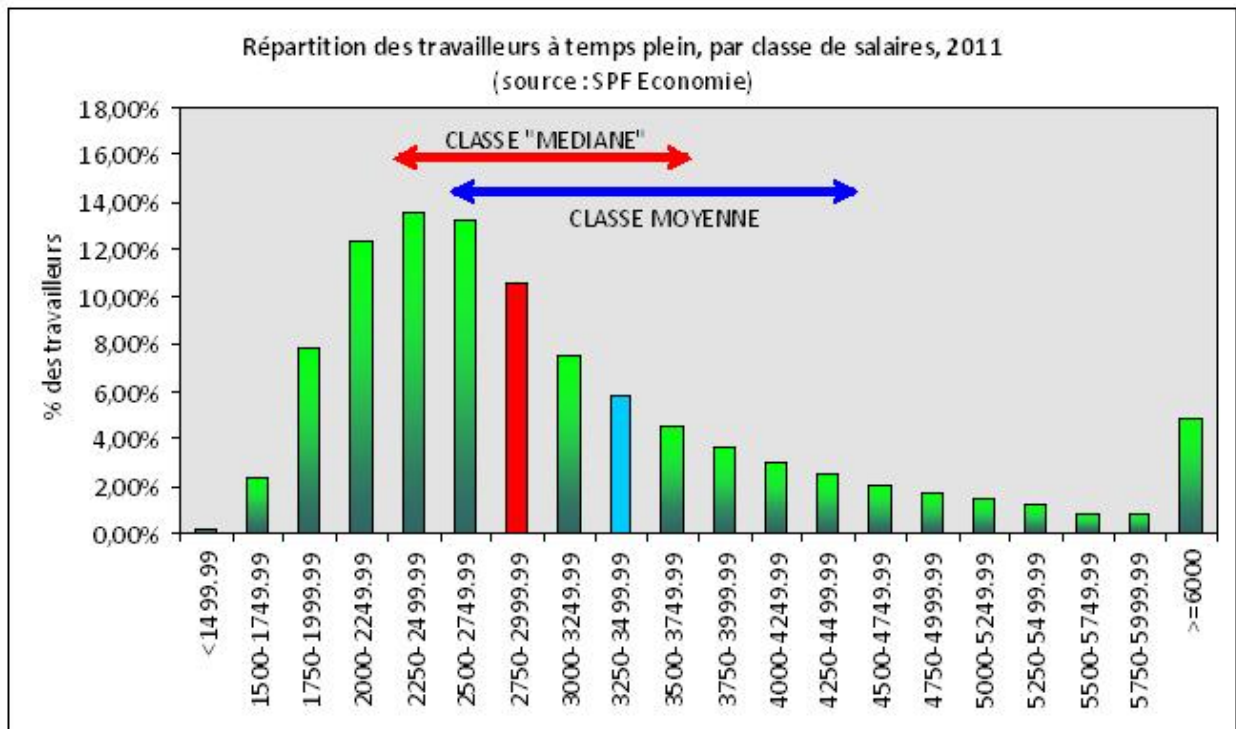


- a) Calcule la moyenne, le mode, la médiane et les quartiles.
- b) Parmi ces valeurs, quelles sont celles qui apportent une information intéressante ?
- c) A l'aide des informations utiles relevées, écrit une dépêche (article court) de trois phrases maximum qui pourrait paraître dans la presse.

2. Question de classes

Un article d'une revue (La Revue Nouvelle) tente de définir la notion de classe moyenne à partir du revenu mensuel brut des travailleurs. Cette notion de classe moyenne est en effet régulièrement utilisée dans les débats politique sans que personne ne la définisse précisément.

Le graphique ci-dessous illustre le propos de l'article.



En recréant un tableau de calcul si nécessaire, réponds aux questions suivantes :

- Selon le graphe du journaliste, défini la classe moyenne et la classe médiane.
- Quels sont approximativement les salaires médian et moyen ?
- Imagine comment le journaliste a pu obtenir sa définition de classes moyenne et médiane.