FÉDÉRATION WALLONIE-BRUXELLES
ENSEIGNEMENT.BE

ÉVALUATION EXTERNE NON CERTIFICATIVE

5e ANNÉE DE L'ENSEIGNEMENT SECONDAIRE

EENC2022

MATHÉMATIQUES

STATISTIQUE

PISTES DIDACTIQUES

Table des matières

Ce document de **Pistes didactiques** a été élaboré par le groupe de travail chargé de la conception de l'évaluation externe non certificative en mathématiques pour la 5e secondaire (enseignement de transition et de qualification) **:**

Marc ANNOYE, inspecteur ;

Donatienne ANTOINE, enseignante ;

Delphine BATTA, conseillère au soutien et à l'accompagnement ;

Laurent BECK, chargé de mission à la Direction des Standards éducatifs et des Évaluations ;

Christelle CORNELIS, conseillère au soutien et à l'accompagnement ;

Marie-Aline DATH, conseillère au soutien et à l'accompagnement ;

Sébastien DELATTRE, attaché à la Direction des Standards éducatifs et des Évaluations ;

Isabelle DEMONTY, chercheuse au Service d'analyse des Systèmes et des Pratiques d'enseignement (Université de Liège) ;

Colette GENOT, inspectrice ;

Valérie GUERRERO, enseignante ;

Fabien JACQUES, inspecteur ;

Sabine LEVEQUE, enseignante ;

Annick LOOZE, conseillère au soutien et à l'accompagnement ;

Frédéric MASCETTI, inspecteur ;

Marianne PITTOIS, enseignante ;

Fabienne POSTAL, conseillère au soutien et à l'accompagnement ;

Charles SOHET, directeur ;

Murielle ZINQUE, conseillère au soutien et à l'accompagnement.

L'emploi dans le présent document des noms masculins pour les différents titres et fonctions est épicène en vue d'assurer la lisibilité du texte.

INTRODUCTION

Ce document fait suite aux résultats de l'évaluation externe en mathématiques menée en octobre 2022 dans les classes de 5e secondaire. Les premiers résultats ont été diffusés sur le site « enseignement.be ». Cette évaluation n'avait pas pour ambition de mesurer le niveau en mathématiques des jeunes de 5e secondaire mais plutôt de **faire un bilan, en début d'année scolaire, des acquis et des faiblesses des élèves dans le domaine de la statistique descriptive enseignée, en 4e secondaire dans l'ensemble des sections.**

Étant donné la diversité des attendus définis dans les référentiels de compétences terminales, le groupe a réalisé trois épreuves :

- une épreuve d'une durée de 2 x 50 minutes pour les élèves de l'enseignement général, technique et artistique de transition ;
- une épreuve d'une durée de 2 x 50 minutes pour les élèves de l'enseignement technique et artistique de qualification, en ciblant les attendus relatifs à l'orientation « Mathématiques actives dans la formation qualifiante » (2 périodes par semaine) ;
- une épreuve d'une durée de 50 minutes pour les élèves de l'enseignement professionnel.

Chaque épreuve comportait une série de questions décomposées en items : cela a permis aux élèves d'exploiter des diagrammes sous différents angles, allant d'une lecture directe d'informations jusqu'à une analyse ou même une critique des informations issues de ces représentations. En outre, un certain nombre de processus à enseigner étaient communs à deux voire aux trois sections : les items ciblant ces processus apparaissaient donc dans différentes épreuves. D'autres processus sont davantage spécifiques à une section, c'est la raison pour laquelle certains items n'apparaissaient que dans une seule épreuve.

Le document se structure en trois parties.

Il débute par une synthèse des constats qui se dégagent de l'analyse des résultats des élèves, en début de cinquième secondaire. Cette synthèse est illustrée par des exemples de questions, auxquelles les enseignants pourront se référer pour identifier la mesure dans laquelle les constats posés à l'échelle de l'ensemble des élèves se retrouvent également au niveau de leur(s) classe(s).

Viennent ensuite deux pistes didactiques visant à travailler des éléments qui posent problèmes aux élèves, en début de 5e secondaire : la première piste revient sur les notions de moyenne arithmétique (nommée « moyenne » dans la suite du document) et médiane et la seconde piste explore l'écart-type. Ces pistes s'adressent aux enseignants de 4e secondaire, puisque le contenu évalué est à enseigner à ce niveau. Les enseignants de 5e secondaire sont également concernés dans la mesure où ils pourront trouver des idées pour réactiver les acquis des élèves, travaillés en 4e secondaire et utiles pour la suite des apprentissages statistiques de 5e secondaire. Ces pistes didactiques sont accompagnées de huit activités exploitables directement en classe, avec les élèves.

PRINCIPAUX RÉSULTATS DE L'ÉVALUATION

1. LES RÉSULTATS GLOBAUX

Afin d'analyser les résultats, les données d'un échantillon représentatif des élèves de chaque filière ont été récoltées et analysées. Le tableau ci-dessous reprend l'ensemble des scores globaux ainsi que les scores par dimension de processus pour l'enseignement général, technique et artistique de transition et de qualification. Les sous-scores par dimension de processus n'ont pu être calculés pour l'enseignement professionnel, en raison du nombre trop faible d'items par dimension dans cette section (il n'y avait en effet qu'une heure de test prévue dans l'épreuve destinée à ces élèves).

		En G (N= 2602)[1]	**En TT** (N=1466)	**En TQ** (N=1814)	**En P** (N=1219)
Score global		**56%**	**49%**	**50%**	**44%**
Scores par dimension de processus évaluée	Connaitre	55%	49%	48%	-
	Appliquer	51%	44%	50%	-
	Transférer	65%	58%	53%	-

Les scores globaux sont compris entre 44% et 56% selon les filières. En outre, pour la première épreuve (destinée aux élèves de l'enseignement général et technique de transition), on constate des différences selon la dimension des processus évaluée : les questions portant sur les processus « transférer » sont mieux réussies que les autres. A l'inverse, les questions portant sur les processus « appliquer » sont en moyenne moins bien réussies.

Une analyse des résultats selon différentes caractéristiques d'élèves a également été réalisée. Les résultats globaux ont été comparés selon le sexe de l'élève, selon le nombre d'heures de mathématiques en 5e secondaire et selon le retard scolaire. Si, dans la filière professionnelle, aucune différence n'est significative, la situation est plus contrastée dans les autres filières : les garçons ont un score moyen légèrement supérieur aux filles (les différences sont comprises entre 3 et 6 points de pourcentage selon les filières). Le nombre d'heures de mathématiques ainsi que le fait de ne pas avoir connu le redoublement ont un impact sur les résultats des élèves, particulièrement dans la filière d'enseignement général (dans cette filière, on observe un écart de 9 points de pourcentage sur le résultat global pour le redoublement et un écart de 16 points de pourcentage selon que les élèves ont 6 heures ou 4 heures de mathématiques par semaine). Enfin, plusieurs écoles ont opté pour une passation électronique de l'épreuve : la plupart de celles-ci organisent un enseignement de transition, ce qui permet d'aller plus loin dans l'analyse des résultats dans cette filière. Au total, 161 élèves de cette filière ont passé le test sur ordinateur et les résultats de ces élèves sont tout à fait équivalents à ceux des élèves qui ont été soumis à un format « papier-crayon ».

[11] Nombre d'élèves sélectionnés pour faire partie de l'échantillon. Le nombre plus important d'élèves dans l'enseignement général s'explique par le fait que les classes sont plus peuplées dans cette filière comparativement aux autres.

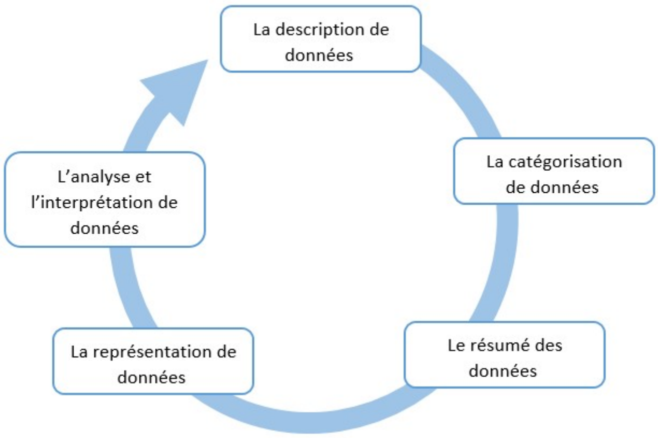
2. L'ANALYSE D'ITEMS

Afin d'affiner les constats, une analyse plus systématique des résultats obtenus à chaque question a été réalisée. Dans cette section, nous synthétisons les constats issus de cette analyse par question et terminons par les grandes orientations qui sont développées dans ces pistes didactiques.

Décrire des données en prélevant une information correcte issue d'un diagramme ou d'un tableau de nombres, catégoriser des données, les résumer au travers d'indicateurs de position ou de dispersion, les représenter ou les interpréter : où se situent les forces et les faiblesses des élèves dans les différentes filières ?

Dans une majorité des classes, la statistique descriptive n'avait pas été revue avant le test, c'est la raison pour laquelle le groupe de travail chargé de la conception des épreuves a privilégié des questions portant sur la compréhension des concepts, sur l'analyse et l'interprétation de données, plutôt que sur la restitution de connaissances, que l'on serait bien entendu en droit d'attendre si la matière venait d'être apprise en classe.

Dans les référentiels de compétences décrivant les attendus au terme de la 4e secondaire, il apparait qu'une réflexion de nature statistique est à développer auprès des élèves, quelle que soit la filière fréquentée. Cinq démarches structurent cette réflexion (Musjadin, Muzaki, Abidin, & Ariyanti, 2020) :



Nous proposons de synthétiser les résultats des épreuves à la lumière de ces cinq types de démarches. Ces constats sont illustrés par des analyses de quelques items issus des différentes épreuves. Cette analyse complète les rapports détaillés disponibles à l'adresse suivante :

www.enseignement.be/evaluationsexternes

Ces derniers envisagent quant à eux une analyse par filière en présentant les scores globaux, la dispersion des résultats pour l'ensemble des élèves et les scores par item.

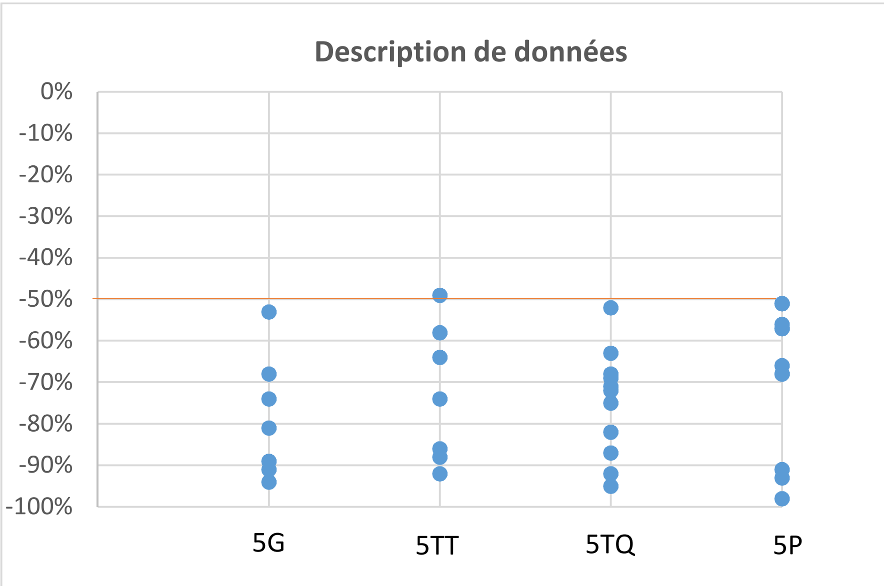
A. LA DESCRIPTION DE DONNÉES

La statistique descriptive implique notamment la capacité à prélever correctement une information explicite issue d'un diagramme ou d'un tableau de nombres. Certains de ces supports sont très facilement accessibles à la lecture directe et immédiate (c'est le cas par exemple, si les effectifs ou fréquences sont indiquées directement au-dessus ou à côté des bâtonnets) alors que d'autres sont moins transparents notamment parce que leur lecture nécessite la prise en compte de connaissances particulières (par exemple la boite à moustache). On retrouve, dans chaque référentiel, des attendus liés à la description de données, comme le montre le tableau suivant.

La description des données dans les référentiels de compétences des différentes filières UAA de statistique, 4e année secondaire	
G/TT	Décrire les informations graphiques et numériques qui peuvent être associées aux différents types de caractères statistiques
	Extraire une information de graphiques et de tableaux statistiques
TQ	Lire les informations fournies par une représentation graphique liée à un ensemble de données statistiques
	Décrire les informations graphiques et numériques qui peuvent être associées aux différents types de caractères statistiques
	Extraire des informations d'une représentation graphique de données statistiques
P	Lire les informations fournies par une représentation graphique liée à un ensemble de données statistiques
	Extraire des informations d'une représentation graphique de données statistiques

Dans chacune des épreuves, certaines questions visaient à voir si les élèves parvenaient à dégager des informations explicites en lisant différents types de supports : des représentations graphiques, telles que l'on peut les trouver dans les médias par exemple, ou des diagrammes statistiques plus formels (tels que la boite à moustache ou l'histogramme dans l'enseignement de transition, les diagrammes circulaires ou en bâtonnets dans l'enseignement qualifiant).

Le graphique suivant présente les pourcentages de réussite des items portant sur la description des données, dans chacune des filières.



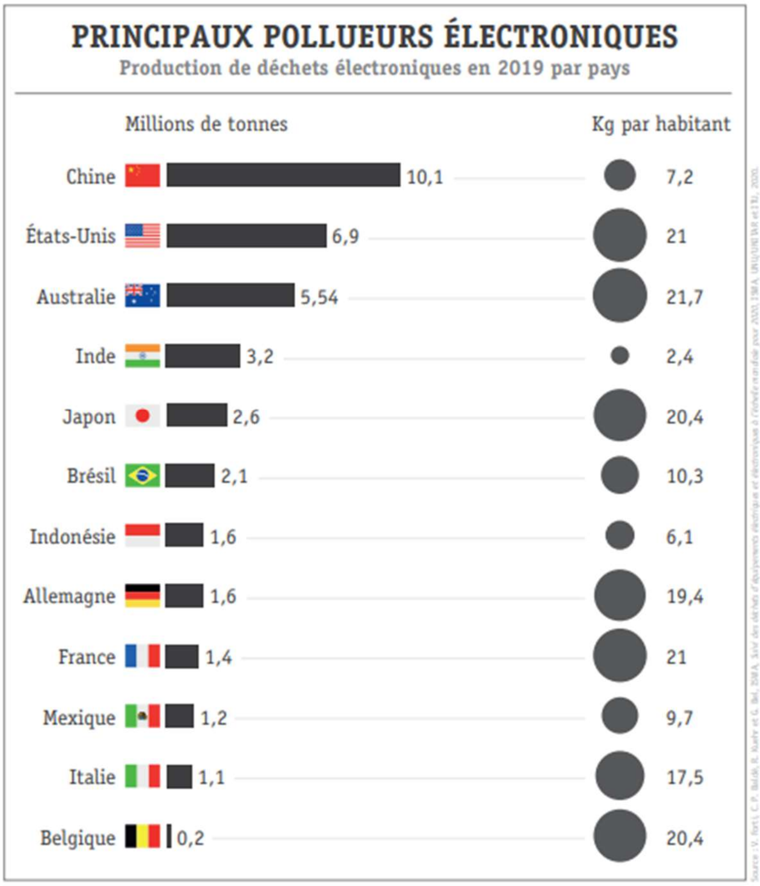
C'est dans le domaine de la description des données que les résultats des élèves des différentes filières sont les meilleurs. Quelle que soit la filière considérée, la plupart des élèves, mêmes les moins performants, ont des acquis dans le domaine de la description de données, particulièrement lorsque ces supports fournissent des indices visuels tels que l'affichage des données par exemple : ces items obtiennent en effet des scores supérieurs à 75% de réussite. En revanche, lorsque les supports sont moins transparents, qu'il faut par exemple prendre en compte les unités des axes dans l'enseignement professionnel ou les concepts de médiane ou de quartile dans l'enseignement de transition, certains élèves sont mis en difficulté : ceux-ci sont donc en quelque sorte dépendants des indices visuels qui les empêchent de faire des déductions plus élaborées.

Quelques illustrations relatives aux acquis et aux faiblesses des élèves en matière de description de données

Les pourcentages de réussite de la question suivante (présente dans les épreuves de l'enseignement général, technique de transition et de qualification) montrent que les élèves parviennent à relever des informations face à un support dont la lecture est assez simple.

QUESTION 11

L'illustration ci-dessous présente la quantité de déchets électroniques produits en 2019 dans différents pays du monde.



a) Les affirmations suivantes correspondent-elles aux données de cette illustration ?
Pour chaque affirmation, trace une croix dans la colonne qui convient.

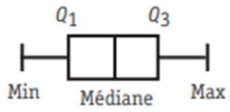
	Oui	Non
La Chine est le pays où l'on produit le plus de déchets électroniques.		
En moyenne, un Belge produit moins de déchets électroniques qu'un habitant des autres pays.		
L'Inde est le pays où l'on produit le moins de déchets électroniques par habitant.		

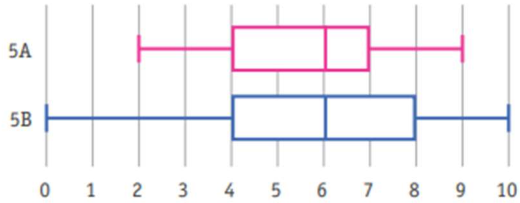
☐ 28
☐ 29
☐ 30

	G	TT	TQ
Item 28	94%	92%	95%
Item 29	89%	89%	75%
Item 30	91%	91%	82%

D'autres questions sont plus problématiques.

- La lecture d'une boite à moustache (qui concerne les élèves de transition) est moins accessible aux élèves.





À quelle(s) classe(s) correspondent les interprétations suivantes ?

Pour chaque ligne, fais une croix dans la case qui convient.

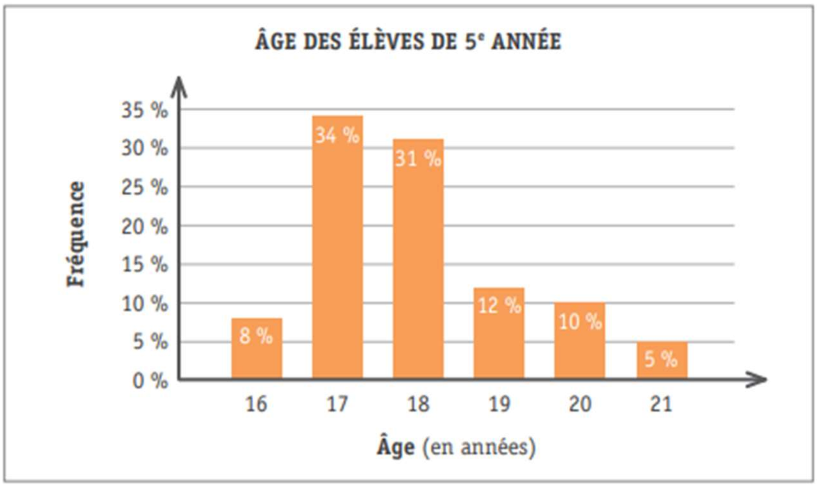
	Ni en 5A, ni en 5B	Uniquement en 5A	Uniquement en 5B	En 5A et en 5B	
Au moins 50 % des élèves ont une note inférieure ou égale à 6.					☐ 47
25 % des élèves ont une note entre 6 et 7.					☐ 48
La note la plus basse obtenue par un élève est de 2.					☐ 49
25 % des élèves ont une note de 9.					☐ 50

	G	TT
Item 47	68%	58%
Item 48	53%	49%
Item 49	81%	74%
Item 50	74%	64%

- Un diagramme de fréquence où on demande d'aller au-delà d'une lecture directe est également plus complexe pour les élèves de l'enseignement de qualification.

QUESTION 7

Dans une école, une enquête a été menée sur l'âge des 102 élèves de 5e année.



b) Les affirmations suivantes correspondent-elles aux données fournies dans le diagramme ? **Pour chaque affirmation**, trace une croix dans la case qui convient.

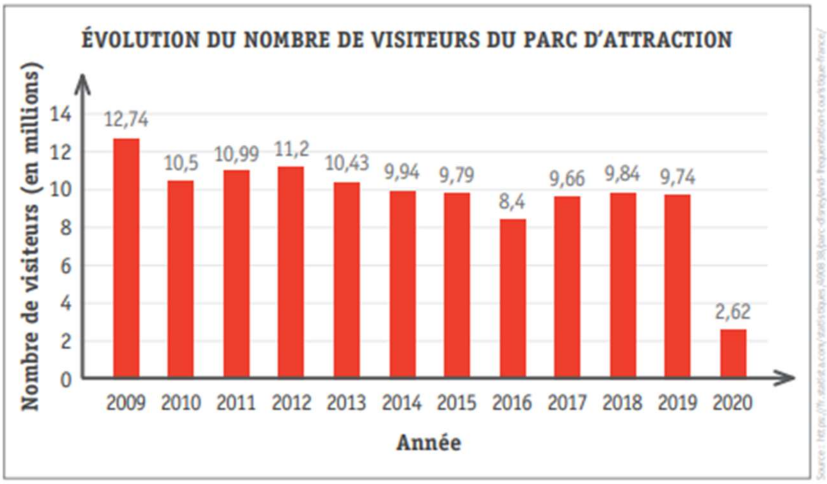
	Oui	Non	Les données ne permettent pas de le dire	
Plus de la moitié des élèves de 5e année de l'école ont plus de 17 ans.				☐ 24
Dans cette école, 31 % des élèves de 5e année sont âgés de 18 ans ou plus.				☐ 25

	TQ
Item 24	63%
Item 25	52%

- Dans l'enseignement professionnel, le contraste de réussite des deux items ci-dessous, portant sur le même support, montre également que, lorsqu'il faut se détacher des indices visuels, l'analyse est plus complexe.

QUESTION 2

Le diagramme suivant présente le nombre (en millions) de touristes ayant visité un parc d'attraction de 2009 à 2020.



a) En quelle année le parc a-t-il été le plus fréquenté ? ______________ ☐ 6

b) Combien de visiteurs ont fréquenté le parc en 2016 ? ______________ ☐ 7

	P
Item 6	98%
Item 7	66%

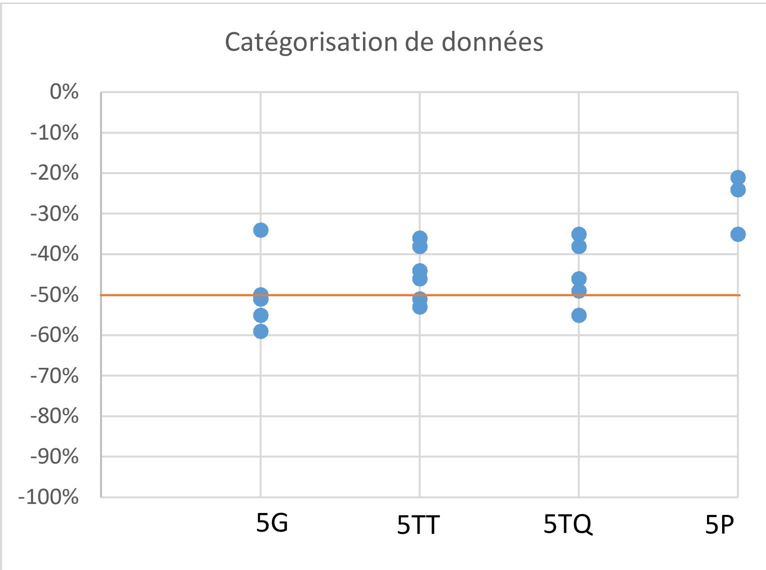
B. La catégorisation de données

Face à un ensemble de données présentées dans un tableau ou simplement évoquées, il est important de se questionner sur la nature des données à traiter : s'agit-il de données quantitatives discrètes ou continues ? De données qualitatives ? Lorsque seul un échantillon est envisagé, à quelle population plus générale se réfère-t-il ? À nouveau, cette thématique fait partie des attendus de la 4e année secondaire, toutes filières confondues, comme le montrent les informations présentées dans le tableau suivant.

La catégorisation de données dans les référentiels de compétences des différentes filièresUAA de statistique, 4e année secondaire	
G/TT	Expliquer le vocabulaire statistique
	Identifier les différents types de caractères statistiques
	Organiser des informations
TQ	Expliquer en situation le vocabulaire caractérisant un ensemble de données statistiques
	Identifier les différents types de variables statistiques
P	Expliquer en situation le vocabulaire caractérisant un ensemble de données statistiques

La catégorisation est investiguée dans les trois épreuves. Dans l'enseignement professionnel et technique de qualification, l'analyse se limite à distinguer les données qualitatives des données quantitatives. Dans l'enseignement de transition, la distinction entre variables quantitatives discrètes et continues est approfondie, de même que la différence entre population et échantillon.

Le graphique suivant présente les pourcentages de réussite des items portant sur la catégorisation de données dans chacune des filières.



Identifier la variable ou le type – qualitatif ou quantitatif – de variable étudiée dans l'enseignement de qualification, affiner l'analyse vers la distinction entre variable quantitative discrète et continue dans l'enseignement de transition sont des thématiques moyennement maitrisées : la plupart des résultats sont en compris entre 30 et 60% de réussite dans l'enseignement général, technique et artistique de transition et de qualification. Les difficultés des élèves du professionnel face à ces items semblent encore bien plus importantes, puisque ces questions sont réussies par 35% des élèves, au maximum.

Ces résultats s'expliquent sans doute en partie par le fait que ces questions mobilisent un vocabulaire spécifique qui n'est pas réellement maitrisé par les élèves, en début de 5e année secondaire : les réponses

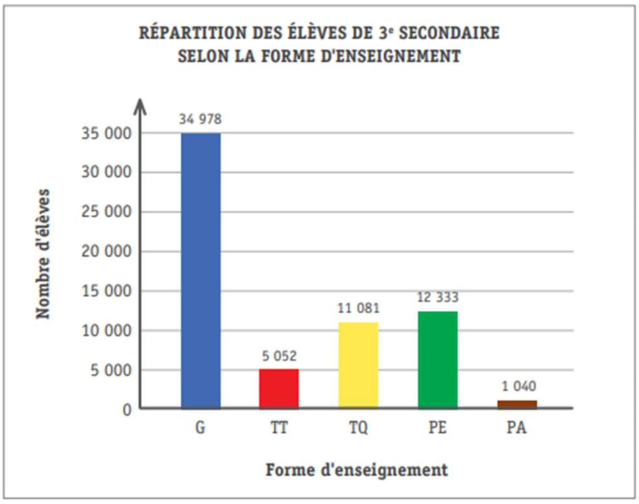
erronées observées lors du prétest de l'épreuve font par exemple apparaitre le fait que, pour un certain nombre d'élèves, toute donnée représentée par des nombres est quantitative, le terme « quantitative » faisant, dans le langage courant, référence à des données numériques. La réflexion sur la nature même des données à traiter n'est donc pas un élément qui retient l'attention des élèves, lorsqu'ils répondent à ces questions. Or, c'est pourtant un élément crucial pour identifier les types de traitements statistiques ou de représentations graphiques que l'on pourra réaliser.**

Quelques illustrations relatives aux acquis et aux faiblesses des élèves en matière de catégorisation de données :

- La question suivante, soumise aux élèves des filières de transition et de technique/artistique de qualification montre les difficultés des élèves dans le domaine de la catégorisation de données.

QUESTION 2

Le diagramme ci-dessous présente la répartition des 64 484 élèves de 3ᵉ année de l'enseignement secondaire en 2019-2020 en Fédération Wallonie-Bruxelles.



Forme d'enseignement	
Général	G
Technique de transition	TT
Technique de qualification	TQ
Professionnel de plein exercice	PE
Professionnel en alternance	PA

a) Quelle est la population ? ☐ 6

____

____

	G	TT	TQ
Item 6	34%	36%	38%
Item 7	59%	53%	49%

b) Coche la variable (ou le caractère). ☐ 7

- ☐ Le nombre d'élèves de 3ᵉ secondaire.
- ☐ La Fédération Wallonie-Bruxelles.
- ☐ L'année 2019-2020.
- ☐ La forme d'enseignement.

- Cette seconde question, soumise aux élèves de l'enseignement technique et artistique de qualification et de l'enseignement professionnel, pose également problème à une majorité d'entre eux :

QUESTION 8

Le tableau suivant présente l'état civil des Belges au 1er janvier 2021.

État civil	Célibataire	Marié	Veuf	Divorcé	Total
Nombre de personnes	5 697 008	4 100 796	651 942	1 071 492	11 521 238

Source : Statbel (Direction générale Statistique - Statistics Belgium)

a) Quelle est la population ? ______________________ 25

b) Quelle est la variable (ou le caractère) ? ______________________ 26

	TQ	P
Item 25	35%	11%
Item 26	55%	24%

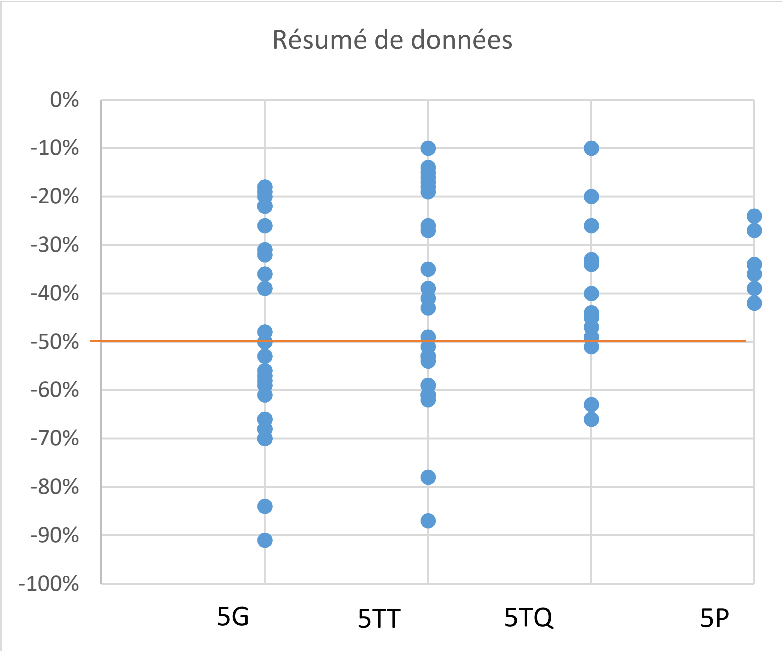
C. Le résumé de données

Les indicateurs de position (tels que la moyenne, le mode ou la médiane) et de dispersion (l'étendue ou l'écart-type par exemple) contribuent à résumer les données à traiter. Ils apparaissent dans les attendus des différentes filières, comme le montrent les informations présentées dans le tableau suivant :

Le résumé de données dans les référentiels de compétences des différentes filières UAA de statistique, 4e année secondaire	
G/TT	Calculer ou estimer les indicateurs de position
	Positionner les indicateurs de position sur un graphique
	Calculer ou estimer les indicateurs de dispersion
	Positionner les indicateurs de dispersion sur un graphique
	Choisir une valeur centrale, un indice de dispersion pour étudier une situation
TQ	Calculer des valeurs caractéristiques d'un ensemble de données statistiques
	Traiter des données statistiques en utilisant l'outil informatique (tableur)
P	Calculer des valeurs caractéristiques d'un ensemble de données statistiques

En référence aux attendus relatifs à chacune des filières, le mode, la médiane et la moyenne sont questionnés dans les épreuves, dans leurs aspects conceptuels (sens des notions) et calculatoires (formules). Dans l'enseignement de transition, l'étendue et l'écart-type sont également envisagés. Toutefois, peu de questions portent sur ces indicateurs de dispersion, vu la volonté du groupe de travail de respecter la réalité des classes en se basant sur les essentiels et balises diagnostiques (thématiques apparaissant en priorité 1) pour l'année scolaire 2021-2022, à la suite de la crise sanitaire.

Le graphique suivant présente les pourcentages de réussite des items portant sur le résumé de données, dans chacune des filières.



En matière d'indicateurs de position, l**a notion de médiane est à revoir dans toutes les filières** : quelle que soit l'épreuve envisagée, les questions portant sur la médiane figurent parmi les moins bien réussies.
Pour le reste, les tendances sont assez différentes selon les filières :

- **Dans la filière générale**, les élèves parviennent globalement à donner sens et à calculer une moyenne pondérée ou non. La plupart sont en revanche mis en difficulté lorsqu'il s'agit d'estimer la moyenne de données présentées en classes.
- **Dans la filière technique de transition et de qualification,** les élèves sont familiers avec la formule de calcul de la moyenne de données brutes. En revanche, donner du sens à celle-ci ou calculer une moyenne pondérée doit encore faire l'objet d'un apprentissage soutenu.
- **Dans la filière professionnelle,** le calcul d'une moyenne simple n'est pas à la portée d'une majorité. En outre, les questions portant sur la compréhension de la moyenne sont complexes pour une majorité d'élèves, puisque la réussite est systématiquement sous la barre des 50%.

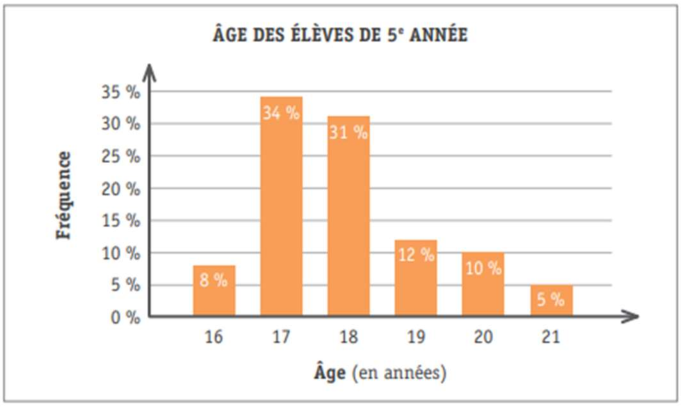
En transition, les élèves ont beaucoup plus d'aisance à utiliser des indicateurs de position que ceux de dispersion (écart-type et étendue). Or, une analyse conjointe de ces deux types d'indicateurs est nécessaire pour analyser un ensemble de données. En outre, l'avis des enseignants de 5e secondaire faisant partie de l'échantillon confirme qu'elle constitue un prérequis essentiel pour aborder la statistique à deux variables.

Quelques illustrations relatives aux acquis et aux faiblesses des élèves en matière de résumé de données :

- Les résultats de la question suivante, présente dans les épreuves de l'enseignement général, technique de transition et de qualification montrent le contraste de réussite des élèves, selon qu'on les interroge sur le mode ou la médiane.

QUESTION 7

Dans une école, une enquête a été menée sur l'âge des 102 élèves de 5e année.



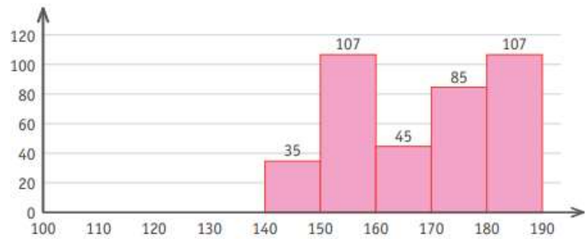
a) Détermine le mode et la médiane de la variable (l'âge).

- le mode est : ______________ ☐ 16
- la médiane est : ______________ ☐ 17

	G	TT	TQ
Item 16	66%	59%	40%
Item 17	32%	27%	20%

- La question suivante, demandant aux élèves de l'enseignement de transition de reconnaitre le calcul correspondant à la moyenne (avec prise en compte des centres des classes) est très mal réussie :

QUESTION 4



b) **Coche le calcul permettant de déterminer la moyenne.** ☐ 11

☐ $\frac{35 + 107 + 45 + 85 + 107}{5}$

☐ $\frac{35 \cdot 140 + 107 \cdot 150 + 45 \cdot 160 + 85 \cdot 170 + 107 \cdot 180}{35 + 107 + 45 + 85 + 107}$

☐ $\frac{35 \cdot 145 + 107 \cdot 155 + 45 \cdot 165 + 85 \cdot 175 + 107 \cdot 185}{35 + 107 + 45 + 85 + 107}$

☐ $\frac{35 \cdot 150 + 107 \cdot 160 + 45 \cdot 170 + 85 \cdot 180 + 107 \cdot 190}{35 + 107 + 45 + 85 + 107}$

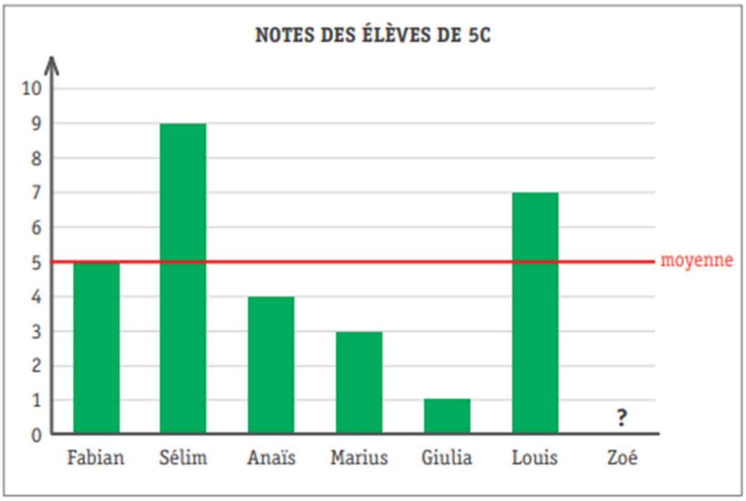
	G	TT
Item 11	31%	19%

- Enfin, les résultats de la question suivante montrent le contraste de réussite, selon qu'on questionne les élèves sur le calcul direct d'une moyenne non pondérée (question b) ou qu'on leur demande une analyse un peu plus approfondie du concept de moyenne (question a).

QUESTION 8

Les 7 élèves de la classe de 5C ont passé un test de français noté sur 10.

Le graphique présente les résultats de la classe, il manque la note obtenue par Zoé.



a) La ligne rouge représente la moyenne des notes des 7 élèves.
Détermine la note obtenue par Zoé et explique ta démarche.

☐ 20

	G	TT	TQ	P
Item 20	57%	49%	26%	/
Item 21	91%	92%	63%	39%

b) Voici les notes des 4 garçons de la classe : 5, 9, 3, 7.
Calcule la moyenne des notes des garçons.

☐ 21

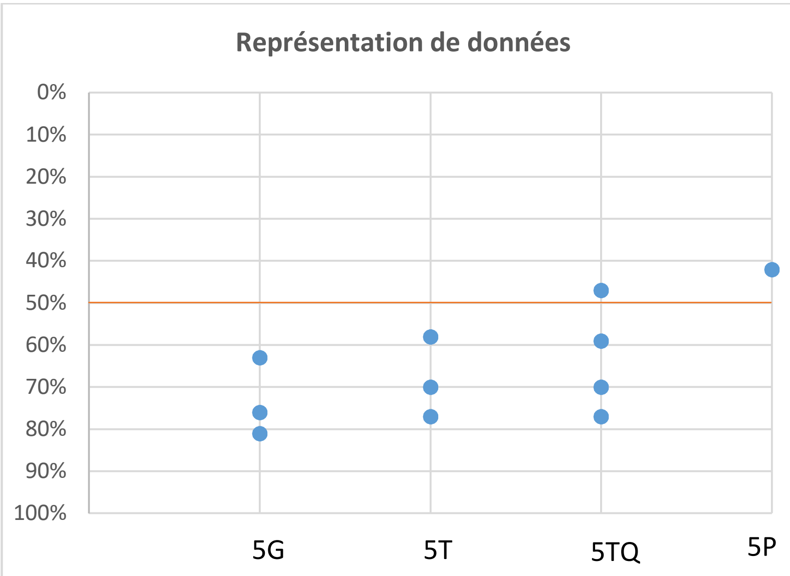
D. La représentation de données

Parvenir à construire des diagrammes statistiques, se questionner sur des supports directement proposés par des logiciels de traitement de données, chercher le support qui permettra le mieux de mettre en évidence une tendance… Tous ces aspects constituent une autre facette de la pensée statistique. On la retrouve dans les attendus des différentes filières comme le montrent les données du tableau ci-dessous :

La représentation de données dans les référentiels de compétences des différentes filières UAA de statistique, 4e année secondaire	
G/TT	Construire différents graphiques statistiques
	Utiliser l'outil informatique dans l'analyse et la présentation des résultats
	Choisir un support graphique pour étudier une situation
TQ	Construire un tableau à partir de données brutes ou recensées
	Construire une représentation graphique liée à un ensemble de données statistiques
P	Construire un tableau à partir de données brutes ou recensées
	Construire une représentation graphique liée à un ensemble de données statistiques

Vu le format des épreuves et le temps imparti pour interroger les élèves, cette dimension a peu été investiguée (entre 1 et 3 items, selon les épreuves) : les élèves ont été invités par exemple à compléter des tableaux de valeurs (calculs de fréquences et de fréquences cumulées), à comprendre pourquoi un diagramme est erroné ou à réfléchir à la manière de présenter des données pour mettre particulièrement l'accent sur un phénomène (dans l'épreuve de G/TT).

Le graphique suivant présente les pourcentages de réussite des items portant sur la représentation de données, dans chacune des filières.



Les quelques informations recueillies concernant la **construction de représentation de données amènent à penser que les élèves sont en bonne voie d'acquisition dans ce domaine**. Outre son intérêt pour la construction des supports graphiques, un travail dans ce domaine pourrait aider les élèves à mieux décrire les données, ou à approfondir la réflexion sur la nature même des variables étudiées (qualitative, quantitative discrète ou continue) qui, comme expliqué précédemment, constitue un élément moins bien maitrisé par les élèves.

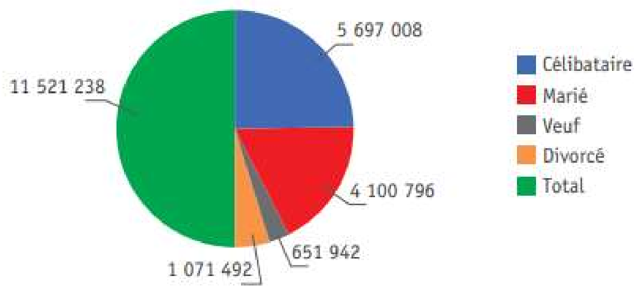
Une illustration relative aux acquis et aux faiblesses des élèves en matière de représentation de données :

La question suivante, présente dans les épreuves de technique de qualification et de professionnelle, montre qu'un nombre non négligeable d'élèves parviennent à trouver une erreur que l'on peut facilement commettre lorsqu'on génère un support graphique au départ d'un logiciel, tel qu'un tableau Excel par exemple.

État civil	Célibataire	Marié	Veuf	Divorcé	Total
Nombre de personnes	5 697 008	4 100 796	651 942	1 071 492	11 521 238

c) Un élève a représenté ces données à l'aide du diagramme circulaire suivant. Son diagramme est faux. 27

ÉTAT CIVIL DES BELGES (AU 1er JANVIER 2021)



Explique l'erreur qu'il a commise.

	TQ	P
Item 27	70%	42%

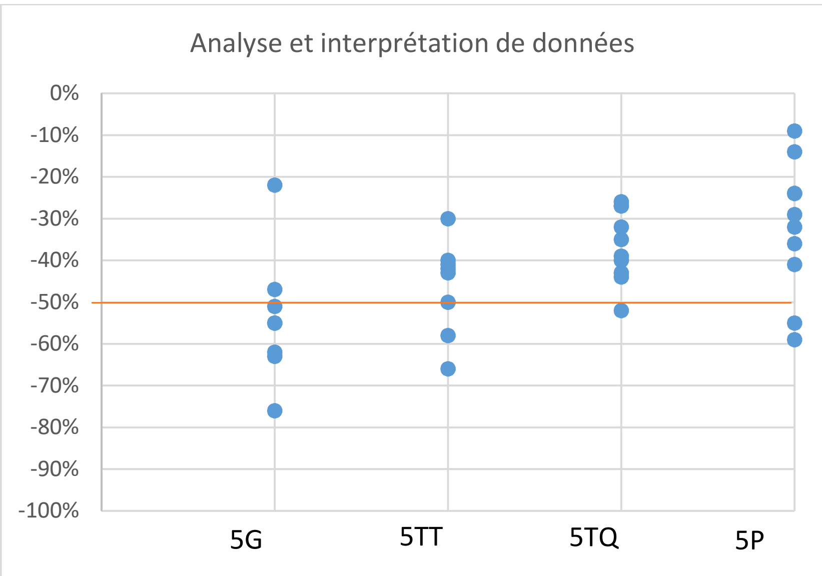
E. L'ANALYSE ET L'INTERPRÉTATION DE DONNÉES

Le dernier aspect envisagé dans les différentes épreuves vise les éléments suivants : identifier des tendances, comprendre des stratégies permettant de mettre en avant certaines données plutôt que d'autres, interpréter des données. Les référentiels des différentes filières d'enseignement insistent particulièrement sur ce thème, comme le montrent les donnéees du tableau suivant :

L'analyse et l'interprétation de données dans les référentiels de compétences des différentes filières - UAA de statistique, 4e année	
G/TT	Décoder les informations statistiques issues de divers contextes
	Développer l'esprit critique
	Synthétiser des informations
	Critiquer des informations graphiques, numériques, textuelles, …
	Commenter des informations fournies sur un même sujet par différents supports
	Interpréter un résultat obtenu en lien avec le caractère étudié et le contexte
TQ	Interpréter en contexte les valeurs caractéristiques d'un ensemble de données statistiques
	Commenter des représentations graphiques liées à un ensemble de données statistiques
	Commenter l'intérêt et les limites d'une étude statistique
P	Interpréter en contexte les valeurs caractéristiques d'un ensemble de données statistiques
	Calculer des pourcentages
	Comparer des rapports en termes des pourcentages
	Calculer des pourcentages successifs

Dans les épreuves, une série d'items invite les élèves à se questionner sur la fiabilité de données, à décrire des tendances générales en mobilisant par exemple la notion de pourcentages ou à se positionner sur des interprétations erronées de données. Dans la filière professionnelle, la notion de pourcentage a été questionnée à de nombreuses reprises (et pas uniquement dans des contextes statistiques), car elle fait partie des apprentissages de 4e secondaire définis dans l'UAA de statistique.

Le graphique suivant présente les pourcentages de réussite des items portant sur l'analyse et l'interprétation de données, dans chacune des filières.



En matière d'analyse et d'interprétation de données, les constats sont variables selon les filières.

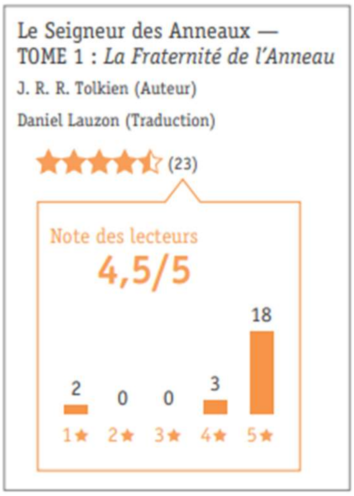
- **Dans l'enseignement général, les compétences des élèves sont en construction** dans le domaine de l'analyse et de l'interprétation : prendre du recul sur l'analyse de données ou développer leur esprit critique dans l'exploitation de données sont des thématiques partiellement maitrisées, comme en attestent les résultats souvent compris entre 50 et 75%. Cependant, deux questions ont été particulièrement mal réussies. Elles invitaient les élèves à s'interroger sur des données présentées dans les médias : l'une leur demandait de mettre en évidence la logique de construction d'un diagramme permettant d'accentuer une tendance et l'autre, de réfléchir à la manière de calculer une donnée présente dans un diagramme statistique.
- **Dans la filière technique de transition, Les élèves doivent approfondir leurs compétences** dans ce domaine, car les résultats sont plus mitigés encore (compris entre 30 et 66%) et une majorité des questions sont réussies par moins de la moitié des élèves.
- **Dans la filière technique de qualification et professionnelle, un travail important est à réaliser dans ce domaine** : seuls quelques items obtiennent des pourcentages de réussite supérieurs à 50%. La notion de pourcentage semble particulièrement peu maitrisée dans l'enseignement professionnel.

Quelques illustrations relatives aux acquis et aux faiblesses des élèves en matière d'analyse et d'interprétation de données

La question ci-dessous interroge les élèves sur l'analyse et l'interprétation de données. Les résultats montrent notamment des faiblesses dans les filières technique, artistique et professionnelle, lorsque l'analyse requiert le calcul d'un pourcentage (items 23 et 24).

QUESTION 9

Sur Internet, on trouve des avis donnés par des lecteurs. Voici les résultats pour deux livres : Le premier tome de la saga du Seigneur des Anneaux et le premier tome de la série des aventures de Harry Potter.



	G	TT	TQ	P
Item 23	55%	40%	32%	14%
Item 24	55%	41%	27%	
Item 25	76%	66%	52%	36%

b) Les lecteurs qui ont aimé le livre attribuent 4 ou 5 étoiles. Quel est le pourcentage de lecteurs qui ont aimé le premier tome du Seigneur des Anneaux ? ☐ 23

_______________ %

c) Les lecteurs mécontents attribuent une seule étoile au livre. Quel est le livre pour lequel le pourcentage des lecteurs mécontents est le plus important ? ☐ 24

Justifie ton choix par un calcul.

____

____

d) Une personne estime que les avis des lecteurs du premier tome d'Harry Potter sont plus fiables que ceux du premier tome du Seigneur des Anneaux. Quelle information appuie cette idée ? ☐ 25

____

____

3. LES ORIENTATIONS DES PISTES DIDACTIQUES

Les indicateurs de positon et de dispersion, au cœur des pistes didactiques

À la lumière des éléments détaillés dans l'analyse précédente, le groupe de travail a choisi de structurer les pistes didactiques en deux parties. La première est centrée sur deux indicateurs de position, la moyenne et la médiane. La seconde propose des activités permettant de prendre conscience des limites des indicateurs de position et approfondit la notion d'écart-type dans ses aspects calculatoire et conceptuel.

Ce choix est motivé par plusieurs raisons.

Tout d'abord, les enseignants de 5e secondaire, interrogés dans le cadre de l'évaluation externe non certificative, considèrent à juste titre qu'une maitrise des indicateurs de position et de dispersion constitue un incontournable de la formation statistique des élèves, et ce dans toutes les filières. Pourtant les résultats des élèves à plusieurs questions concernant ces indicateurs sont préoccupants : quelle que soit la filière, des difficultés apparaissent dans la compréhension des indicateurs mais aussi parfois dans leur calcul.

En outre, cette exploitation des indicateurs de position et de dispersion peut permettre des approches qui dépassent largement le calcul en tant que tel des indicateurs. C'est en effet l'option privilégiée dans ces pistes. L'objectif est de proposer des activités qui visent avant tout à faire réfléchir les élèves concernant ces concepts, dans le but de leur permettre de mieux comprendre leurs apports mais aussi leurs limites respectives. Ces pistes ne sont évidemment pas destinées à remplacer des activités que l'on trouve déjà dans la plupart des manuels ou des cours des enseignants mais plutôt d'enrichir la réflexion des élèves, en cherchant à développer, au travers de l'exploitation des indicateurs de position ou de dispersion, une réflexion de nature statistique.

PARTIE 1 MOYENNE ET MÉDIANE

1. CONSTATS ISSUS DES ÉPREUVES

Tant la moyenne que la médiane ont été explorées dans les questions de l'épreuve, en envisageant les aspects calculatoires (maitrise des techniques de calculs) et conceptuels (sens de la moyenne ou de la médiane dans différentes situations). Dans cette section, nous revenons plus en profondeur sur les constats se dégageant des épreuves, en regard de ces deux concepts.

1.1 La moyenne

L'analyse du taux de réussite aux questions portant sur le calcul de la moyenne fait apparaitre des faiblesses dans la maitrise des formules, et ce quelle que soit la filière envisagée.

Différents calculs de la moyenne ont été demandés aux élèves : la moyenne de données brutes dans toutes les sections, la moyenne de données recensées, organisées en classes (en G et TT) ou non (en TQ).

Dans l'enseignement professionnel, le calcul d'une moyenne de données brutes n'est pas encore acquis. Dans l'enseignement technique et artistique de transition et de qualification, la moyenne de données brutes est acquise par une majorité d'élèves. En revanche, le calcul d'une moyenne de données recensées n'est pas maitrisé : un nombre important d'élèves ne mutliplient pas les effectifs par les valeurs de la variable dans leurs calculs. Enfin dans l'enseignement général, le calcul d'une moyenne pondérée lorsque les données ne sont pas organisées en classes est assez bien maitrisé. En revanche, lorsque les données sont organisées en classes, beaucoup d'élèves n'utilisent pas le centre des classes pour effectuer le calcul.

L'épreuve ayant été organisée en début de 5e secondaire, il est possible qu'un simple rappel des démarches suffise pour que certains élèves comblent ces lacunes. Toutefois, cette difficulté à retrouver une formule peut également cacher un problème de compréhension plus large du concept même de moyenne.

La maitrise calculatoire n'est donc qu'un aspect des connaissances que doivent développer les élèves autour de la moyenne. Les résultats de l'épreuve montrent que la compréhension du concept de moyenne n'est que partiellement acquise par les élèves.

Plusieurs questions de l'épreuve visaient à voir la mesure dans laquelle les élèves parvenaient à se détacher de la formule de la moyenne, pour s'intéresser à ce qu'elle signifie. Les résultats à ces questions montrent que les élèves rencontrent des difficultés pour donner du sens à la moyenne, comme l'illustrent les questions ci-dessous.

Un exemple de question où la moyenne est vue comme la valeur représentative d'une distribution équivalente et uniforme des données.

QUESTION 12

Une enquête s'intéresse aux salaires de 1 000 étudiants ayant travaillé durant 1 mois à temps plein. La moyenne, le mode et la médiane ont été calculés à partir des données recueillies.

Moyenne	1 235 €
Mode	1 183 €
Médiane	1 215 €

a) Complète chaque phrase suivante avec **le nombre** du tableau qui convient.

- La moitié des étudiants interrogés gagnent au moins ________________ €. ☐ 33
- Le salaire le plus fréquemment perçu vaut ________________ €. ☐ 34
- Si la somme totale des salaires est distribuée de manière égale entre tous les étudiants, alors chaque étudiant reçoit ________________ €. ☐ 35

b) Pour calculer la moyenne, par quel nombre a-t-on divisé la somme des salaires ? ☐ 36

Le dernier item de la question a) permet d'analyser si les élèves ont une compréhension de la moyenne en tant que distribution équitable des données.

La question b) questionne davantage la formule.

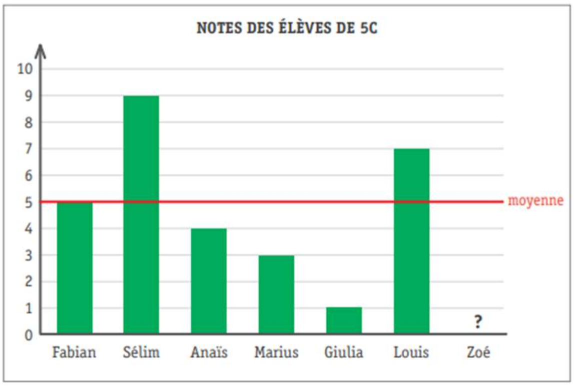
Le contraste entre la réussite aux 2 items est important, quelle que soit la filière d'étude. La notion de division en tant que partage équitable reste donc problématique.

	G	TT	TQ	P
Item 35	70	62	45	22
Item 36	84	78	57	36

Un exemple de question où la moyenne est vue comme un point d'équilibre entre les différentes données

QUESTION 8

Les 7 élèves de la classe de 5C ont passé un test de français noté sur 10.
Le graphique présente les résultats de la classe, il manque la note obtenue par Zoé.



a) La ligne rouge représente la moyenne des notes des 7 élèves. ☐ 20
Détermine la note obtenue par Zoé et explique ta démarche.

b) Voici les notes des 4 garçons de la classe : 5, 9, 3, 7. ☐ 21
Calcule la moyenne des notes des garçons.

La partie a) de cette question demande aux élèves de résoudre un problème assez simple à effectuer si l'élève s'appuie sur la somme des écarts (positifs et négatifs) à la moyenne. La partie b) quant à elle est focalisée sur le calcul en tant que tel.

Le constraste de réussite entre ces deux questions est à nouveau important, quelle que soit la filière envisagée. En outre, les résultats du prétest ont montré que, dans la partie a), moins de 7% des élèves s'appuient explicitement sur les écarts à la moyenne pour déterminer la quantité inconnue.

	G	TT	TQ	P
Item 20	57%	49%	26%	/
Item 21	91%	92%	63%	39%

Une compréhension conceptuelle de la moyenne d'une série de valeurs permet d'aborder différentes propriétés (Batanero, Godino, Vallecillos, 1994) :

- la moyenne est située entre les valeurs extrêmes ;
- elle est influencée par l'ensemble des valeurs ;
- elle ne doit pas nécessairement être une des valeurs de la série ;
- une valeur nulle a une influence sur la moyenne ;
- la somme des écarts (envisagés comme des quantités signées) à la moyenne est nulle.
  En plus de son intérêt pour approcher la moyenne, cette dernière propriété est également utile pour donner sens au calcul de l'écart-type, comme nous le développerons dans la deuxième partie de ces pistes.

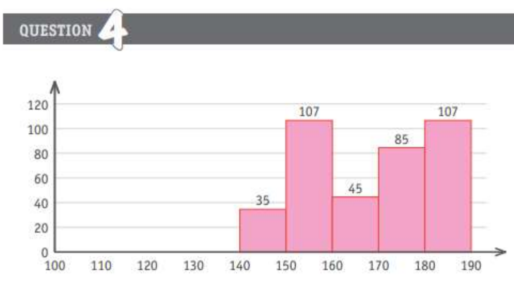
Associer aux opérations le sens de la moyenne est important dans une série de situations : par exemple pour décider de la pertinence d'utiliser la moyenne face à une série de données à analyser ou pour porter un regard critique sur le calcul d'une moyenne effectué par un logiciel. C'est le premier aspect qui sera développé dans les pistes didactiques.

1.2 La médiane

La médiane peut être vue comme un indicateur de tendance centrale plus accessible que la moyenne : en effet, il suffit d'ordonner les données de la série et de dégager celle qui se situe exactement au milieu de celle-ci. **Pourtant, les résultats de l'épreuve montrent que la médiane est très mal maitrisée par les élèves : quelle que soit la filère envisagée, on retrouve, dans les questions les moins bien réussies des épreuves, des calculs élémentaires de médiane.**

Quelles sont les difficultés des élèves dans le domaine ? Les recherchent mettent en évidence qu'en général, **les élèves considèrent que la médiane est « le nombre du milieu ». Ils n'ont pas le réflexe d'ordonner ces données par ordre croissant. De nombreux élèves n'ont pas non plus conscience que lorsqu'il y a un nombre pair de données, deux valeurs sont centrales et qu'il convient donc de calculer la moyenne artihmétique de celles-ci** (Sawojewski & Shaughnessy, 2000).

Les résultats des deux questions ci-dessous sont interpellants. En ce qui concerne la première question, nous avons constaté que, lors du prétest de l'épreuve, plus de la moitié des élèves ont considéré que la classe médiane était la classe [160 ;170[: ces élèves ne se sont pas basés sur les effectifs proprement dit pour déterminer la classe médiane, mais ont sans doute eu en tête la recherche du « nombre du milieu » qui visuellement est interprété comme le bâtonnet central.



a) En utilisant cet histogramme, coche :

la classe médiane

☐ [140 ; 150[
☐ [150 ; 160[
☐ [160 ; 170[
☐ [170 ; 180[
☐ [180 ; 190]

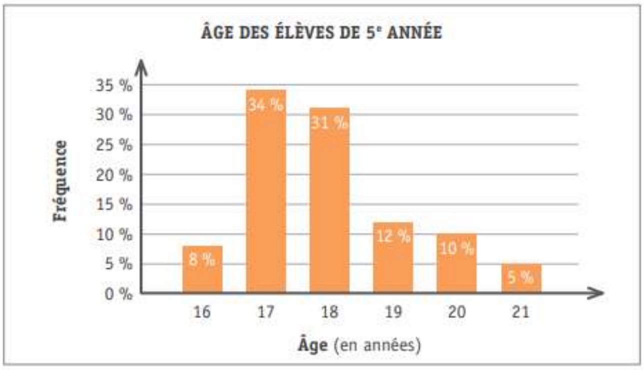
☐ 10

G	TT
20%	18%

Et face à la question suivante, les résultats sont également problématiques dans toutes les filières. À nouveau, nous avons pu constater lors du prétest de l'épreuve que l'erreur la plus fréquente consiste à proposer la réponse 18,5 : celle-ci peut également être considérée comme une recherche du «nombre du milieu ».

QUESTION 6

Dans une école, une enquête a été menée sur l'âge des 102 élèves de 5ᵉ année.



a) Détermine le mode et la médiane de la variable (l'âge).

- la médiane est : ________ ☐ 17

	G	TT	TQ
Item 17	32%	27%	20%

Donner du sens et calculer une médiane dans différentes situations mérite une attention particulière.

Dans les pistes didactiques, nous proposons donc un travail sur la médiane au départ d'une appliquette GeoGebra qui permet aux élèves d'analyser la médiane de différentes séries statistiques. Dans l'activité, les élèves mènent également une réflexion sur les différences conceptuelles entre médiane et moyenne : ils conscientisent le fait que cette dernière implique de réaliser un calcul avec toutes les données de la série, contrairement à la médiane. Cette compréhension est en effet importante pour que les élèves parviennent à choisir un de ces deux indicateurs de tendance centrale, en fonction de la situation.

2. INTENTIONS ET COMMENTAIRES

Les pistes sont regroupées selon trois thématiques :

- la moyenne
- la médiane
- moyenne ou médiane : que choisir ?

Les indicateurs statistiques sont des nombres utilisés pour caractériser un ensemble de données. Les élèves doivent bien comprendre ce que chacun représente afin de les choisir, de les déterminer et de les utiliser de façon appropriée.

2.1 La moyenne

Le concept de moyenne est actuellement à certifier en fin de 6e primaire. Quelle que soit l'année d'enseignement, lorsqu'on manipule le concept de moyenne, il importe de mettre l'accent sur la compréhension plutôt que sur la mémorisation de la définition qui constitue aussi l'algorithme usuel : somme des données divisée par le nombre de données. En effet, de nombreuses études réalisées auprès des élèves de 11 à 16 ans[2] montrent que les élèves ne comprennent pas les propriétés relatives à cet algorithme usuel : par exemple, la moyenne de deux ensembles de données dont l'un comporte 10 données dont la moyenne vaut 3 et l'autre, 20 données dont la moyenne vaut 5, n'est pas égale à 4 ($\frac{3+5}{2}$) comme le pensent de nombreux élèves, mais à 4,3 ($\frac{3 \cdot 10\, +\, 5 \cdot 20}{30}$). Cela entraine des difficultés pour l'utiliser dans différentes situations, par exemple pour calculer une moyenne pondérée. Ces études ont montré que les erreurs des élèves n'étaient pas dues à un manque de connaissance de l'algorithme mais plutôt à une mauvaise compréhension du concept de moyenne.

Le groupe de travail a privilégié deux modèles présents dans la littérature scientifique (Uccellini, 1996) pour guider le choix des activités proposées.

Modèle de partage équitable

« En 2020, un travailleur occupé à temps plein gagnait en moyenne 3.832 euros brut par mois[3]. » Cette information signifie que si la masse salariale restait inchangée et que chaque employé occupé à temps plein percevait le même salaire, celui-ci s'élèverait à 3.832 euros brut par mois.

La moyenne correspond à la valeur résultant d'un **partage équitable** ou plus généralement la valeur représentative d'une distribution équivalente et uniforme, entre les individus d'une population.

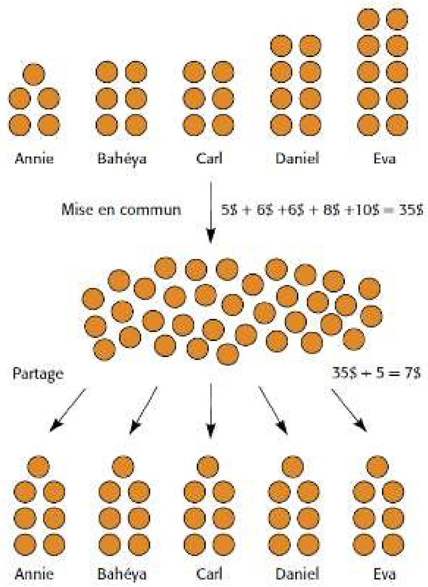
---

[2] Cai, J. (2010). Exploring students' conceptual understanding of the averaging algorithm. School Science and mathematics. Volume 98, issue 2 pp. 93-98.

[3] https://statbel.fgov.be/fr/nouvelles/le-salaire-brut-moyen-en-belgique-seleve-3832-euros-par-mois#:~:text=En%202020%2C%20un%20travailleur%20occup%C3%A9,moyenne%20plus%20que%20les%20hommes

Exemple

Si 5 amis ont récolté respectivement 5 €, 6 €, 6 €, 8 € et 10 € et qu'ils mettent ces montants en commun pour ensuite les partager en parts égales, chacun recevra 7 €.
La moyenne est donc égale à 7 €.



Démarche suivie :

- On met d'abord les montants récoltés en commun : 5 €+6 €+6 €+8 €+10 € = 35 €
- La somme est ensuite partagée en 5 parties égales : 35 € : 5 = 7 €

⇨ Ce modèle donne un sens à l'algorithme usuel puisqu'il démontre l'idée de grouper les montants (somme) puis de partager le résultat en parts égales entre les 5 personnes (division).

Modèle d'équilibre entre les surplus et les manques

Ce modèle est fondé sur l'idée que si, par exemple, un groupe d'élèves dispose d'un nombre moyen de jetons, certains pourraient avoir moins de jetons que la moyenne alors que d'autres pourraient en avoir plus. Cependant, le total de ce que les élèves ont en moins doit être égal au total de ce que les élèves ont en plus.

Surplus ou écarts positifs = différences entre la moyenne et les données qui sont supérieures à la moyenne.
Manques ou écarts négatifs = différences entre la moyenne et les données qui sont inférieures à la moyenne.

⇨ Les écarts positifs (surplus) équilibrent les écarts négatifs (manques).

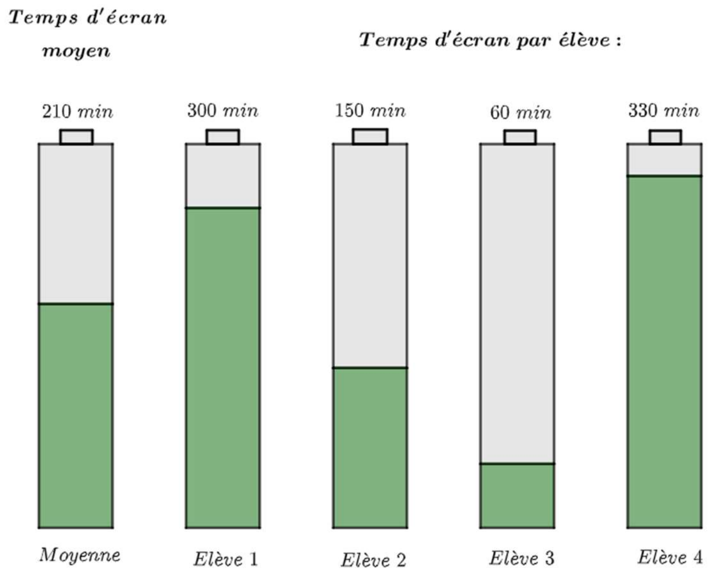
Si un nombre est la moyenne d'une série de données, alors, la somme des écarts entre cette moyenne $(\overline{x})$ et les données (x_i) est nulle.

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

Et réciproquement, si la somme des écarts entre les données d'une série et d'un nombre est égale à 0, alors ce nombre est la moyenne des données.

Exemple

Quatre élèves d'une même classe ont déterminé qu'ils étaient, en moyenne, 3h30 (210 minutes) par jour devant les écrans. Dans cette situation, combien de temps chaque jeune pourrait-il avoir passé par jour devant un écran ?



Pour les élèves de 4e GT, cette propriété peut être déduite de la définition formelle de la moyenne :

Prenons un ensemble constitué de n $(n \in N_0)$ données. La moyenne est notée $\overline{x}$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

$$n \cdot \overline{x} = x_1 + x_2 + x_3 + \cdots + x_n$$

$$0 = (x_1 + x_2 + x_3 + \cdots + x_n) - n \cdot \overline{x}$$

Or, $n \cdot \overline{x} = \overline{x} + \overline{x} + \overline{x} + \cdots + \overline{x}$ (somme de n termes)

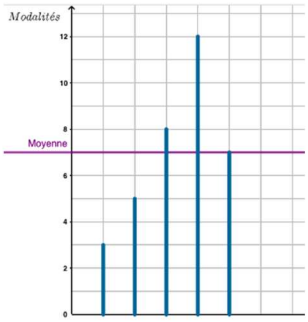
$$0 = (x_1 + x_2 + x_3 + \cdots + x_n) - (\overline{x} + \overline{x} + \overline{x} + \cdots + \overline{x})$$

$$0 = x_1 - \overline{x} + x_2 - \overline{x} + x_3 - \overline{x} + \cdots + x_n - \overline{x}$$

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

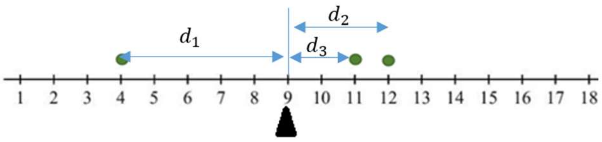
La compréhension et l'appropriation par l'élève de ce modèle d'équilibre entre les surplus et les manques peut s'appuyer sur des supports visuels différents.

- **Approche visuelle 1 :** les élèves travaillent avec un ensemble de données individualisées modélisées par des traits verticaux dont la hauteur correspond à la valeur de la modalité. La moyenne est représentée par une droite horizontale et sa valeur est estimée par l'élève en ajustant la position de la droite pour que les surplus équilibrent les manques.



https://www.geogebra.org/m/yt9ydyrh

- **Approche visuelle 2** : les élèves travaillent sur une représentation physique de la moyenne, envisagée comme le point d'équilibre d'une bascule. Ce modèle visuel les amène d'abord à construire une série statistique dont la moyenne est fixée. Ensuite, ils estiment la moyenne d'une série statistique donnée, en s'appuyant sur les écarts à la moyenne.



$d_1 = d_2 + d_3$

Accéder à la simulation :

https://phet.colorado.edu/fr/simulations/balancing-act

2.2 La médiane

Comme détaillé précédemment, les élèves ont tendance à interpréter la médiane comme le milieu d'une série, le centre de « quelque chose », sans pleinement comprendre ce que représente ce « quelque chose ».

D'un point de vue des mathématiques, derrière cette notion de « valeur centrale », se cachent quatre caractéristiques (Mayen & Diaz, 2010) de la médiane :

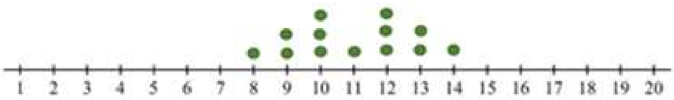
- la valeur qui laisse le même nombre de donnéees avant et après elle ;
- la donnée centrale d'un nombre impair de données ordonnées ;
- la moyenne des deux données centrales d'un nombre pair de données ordonnées ;
- la première valeur de la variable qui correspond à la fréquence cumulative strictement supérieure à la moitié du nombre de données.

La manière de déterminer la médiane n'est donc pas unique et dépend de la façon dont les données sont présentées. C'est un nœud important de difficulté pour les élèves : ils doivent prendre conscience que la médiane se base sur la relation d'ordre entre les données mais également que le concept de valeur centrale est présent, quelle que soit la manière de déterminer la médiane (Schuyten, 1991).

Liste de données ordonnées

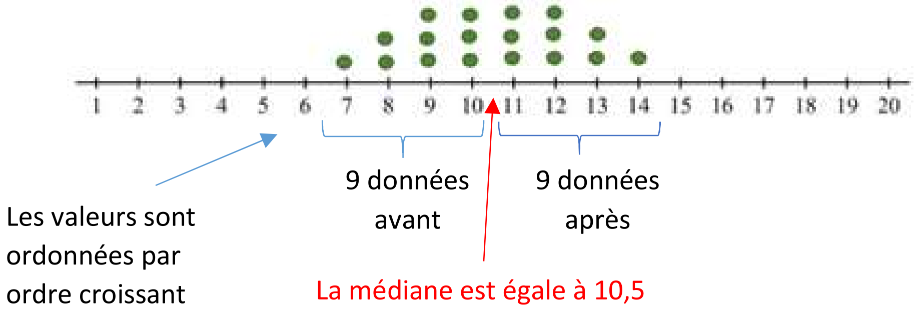
- Si le nombre de données est **impair**, la médiane est toujours parfaitement déterminée :

$\{8, 9, 9, 10, 10,10, 11, 12, 12, 12, 13, 13,14\}$➔ la médiane est 11



- Si le nombre de données est **pair**, on peut définir un intervalle médian à l'intérieur duquel toutes les valeurs vérifient la définition.

$\{7, 8, 8, 9, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 12, 13, 13, 14\}$➔ L'intervalle médian est délimité par les modalités 10 et 11.



Dans cette situation (N=18), la valeur de la médiane est, par convention, le milieu de l'intervalle médian : $\frac{\mathbf{10+11}}{\mathbf{2}} = 10,5$.

Lorsque les données sont présentées sous la forme **d'une liste ordonnée**, l'idée de valeur centrale est assez directe, permettant d'établir le lien entre le concept de médiane et la manière de la déterminer.

Données groupées (tableau de distribution)

Exemple : Distribution groupée du nombre d'enfants par famille

Nbre d'enfants	Nbre de familles	Fréquences	Fréquences cumulées
0	20	0,1	0,1
1	65	0,325	0,425
2	70	0,35	0,775
3	30	0,15	0,925
4	10	0,05	0,975
5	5	0,025	1
	N = 200		

La valeur de la médiane est la valeur de la modalité correspondant à la fréquence cumulée à partir de laquelle 50% de la population sont pris en compte. Les modalités 0 et 1 représentent seulement 42,5% de la population. Il faut donc y ajouter une certaine partie des données ayant pour valeur « 2 » pour atteindre 50 % des effectifs. La médiane, dans ce cas, vaut donc 2.

La notion de valeur centrale est moins visible lorsque les données sont présentées sous la forme d'un tableau recensé que lorsqu'elles apparaissent sous forme de liste : cette notion de tendance centrale passe par une compréhension de la fréquence cumulée.

Données présentées graphiquement (cas d'une variable continue)

Voici la répartition des ouvriers de l'entreprise XY selon le salaire mensuel x_i en euros.

		Centres de classe	Effectifs	Effectifs cumulés
	$[1000;1100[$	1050	26	26
	$[1100;1200[$	1150	33	59
Classe médiane →	$[1200;1300[$	1250	64	123
	$[1300;1400[$	1350	7	130
	$[1400;1500[$	1450	10	140
			N=140 N/2=70	

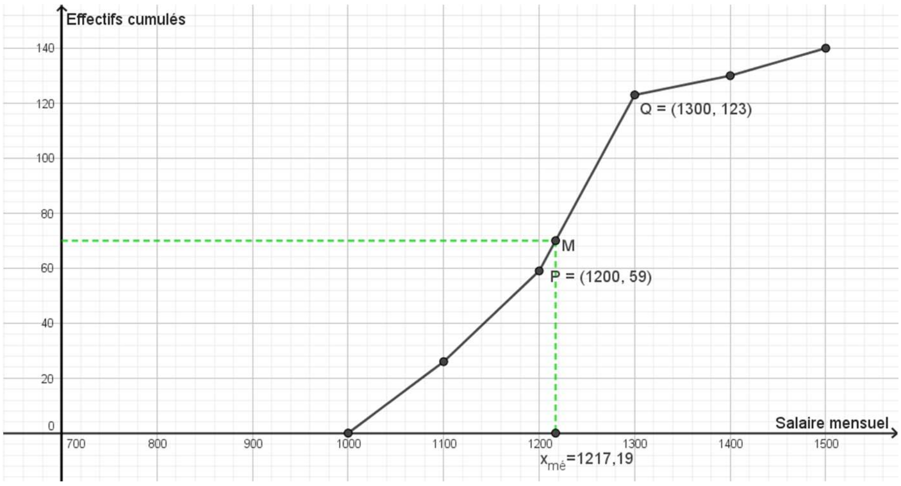
La valeur de la médiane se trouve à l'intérieur de la classe $[1200;1300[$

Comment calculer la valeur exacte de la médiane ?

1) Calculer les effectifs cumulés.
2) Repérer la classe médiane.
3) Sur base du polygone des effectifs cumulés, calculer la valeur de la médiane par interpolation linéaire.

La médiane appartient à la classe $[1200; 1300[$ et sa valeur se calcule par interpolation linéaire car à l'intérieur de cette classe les données sont supposées être réparties de manière uniforme (hypothèse sous-jacente lorsqu'on répartit une série en classes). Cette hypothèse est très importante pour relier la manière de déterminer la médiane à la définition de celle-ci en tant que valeur centrale d'une série statistique.

Polygone des effectifs cumulés



Notons P : (1200 ;59) Q : (1300 ;123) M : ($x_{mé}$;70)

Le coefficient angulaire du segment PQ = Le coefficient angulaire du segment PM

$$\frac{123-59}{1300-1200} = \frac{70-59}{x_{mé}-1200}$$

$$\frac{64}{100} = \frac{11}{x_{mé}-1200}$$

$$x_{mé} = 1217{,}1875$$

Interprétation

- 50% des ouvriers gagnent au plus 1217€, 50% des ouvriers gagnent au moins 1217€.
- 1217€ est le salaire qui partage la population ouvrière de l'entreprise en deux parties égales.

Point d'attention

Dans ce document, notre choix pour calculer la médiane par interpolation linéaire s'est porté sur le coefficient angulaire. D'autres notions peuvent être exploitées : les triangles semblables, la proportionnalité, l'équation de la droite…

ACTIVITÉ 1 : ABORDER LA MOYENNE À L'AIDE D'UNE APPLIQUETTE GEOGEBRA

Cette activité se structure en trois fiches, poursuivant les objectifs suivants :

- Estimer une moyenne en s'appuyant sur les écarts (Fiche 1) ;
- Décrire l'impact sur la moyenne d'un changement de données (Fiche 2) ;
- Créer un ensemble de données ayant une moyenne imposée (Fiche 3).

En ce sens, ces trois fiches contribuent à amener les élèves à se détacher de l'algorithme usuel de calcul de la moyenne pour développer une compréhension plus profonde de ce concept.

La fiche 1 vise à faire émerger les méthodes spontanément utilisées par les élèves pour estimer une moyenne, en vue d'approfondir ces dernières : en analysant des diagrammes en bâtonnets, on met en évidence les écarts (positifs ou négatifs) entre la moyenne et chaque donnée d'une série statistique. Cette propriété est en effet très utile pour développer des stratégies efficaces d'estimation de moyenne.

La fiche 2 amène les élèves à visualiser, à l'aide de l'appliquette GeoGebra, l'impact de changements de données sur la moyenne. L'objectif est d'expliquer, à l'aide d'une réflexion sur les écarts à la moyenne, en quoi la modification dans un sens ou dans l'autre d'une donnée affecte la moyenne.

La fiche 3 propose des exercices où il s'agit de déterminer la donnée manquante de séries statistiques dont la moyenne est connue. À nouveau, une compréhension du concept de moyenne est nécessaire pour résoudre ces problèmes qui autorisent en outre une variété de démarches de résolution.

Quelques exercices complémentaires sont proposés dans la **fiche 4**.

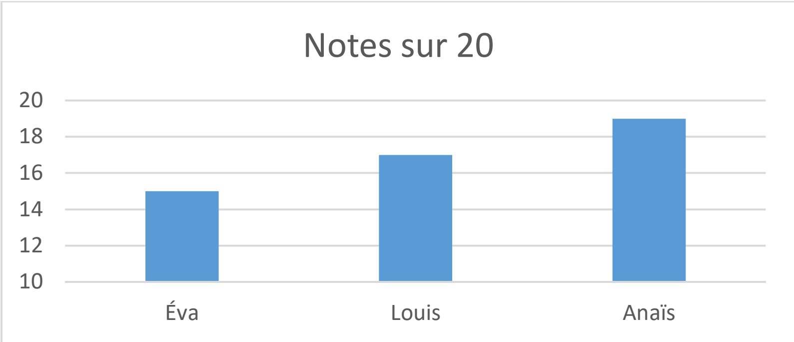
Fiche 1

Estimer la moyenne en s’appuyant sur les écarts

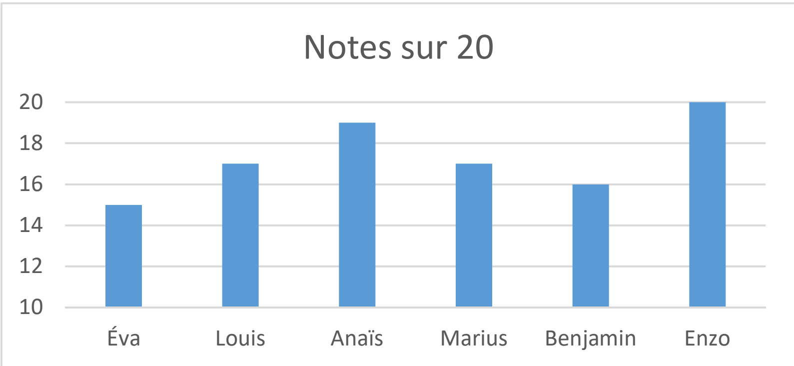
1. Voici les résultats d’élèves obtenu à un test noté sur 20.
   Dans chaque cas,

   a) Estime (sans calcul) la moyenne des notes obtenues.
   b) Vérifie la valeur proposée en utilisant la somme des écarts à la moyenne.
   Tu peux t’aider en représentant sur le diagramme la moyenne, les surplus et les manques.

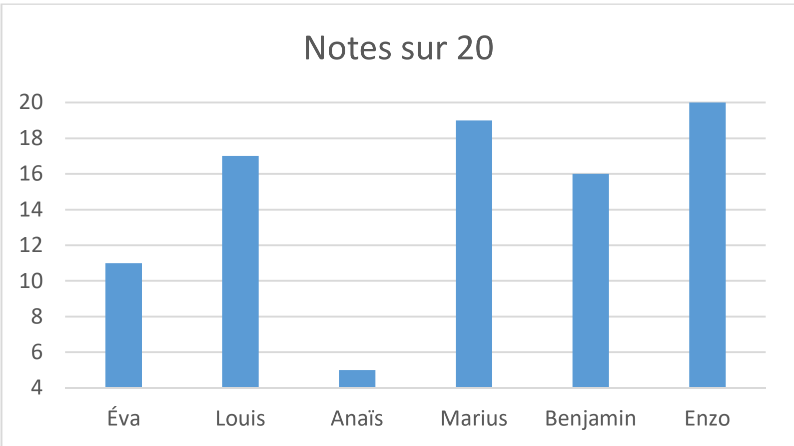
Exemple 1



Exemple 2



Exemple 3



2. Voici les notes sur 20 en mathématique obtenues par 3 élèves.

Dylan	10	11	14
Camille	13	10	16
Sabrina	9	10	14

a) Dylan pense avoir une moyenne de 10 sur 20.
Sans utiliser la formule de la moyenne, explique pourquoi ce n'est pas le cas.

b) Camille pense avoir une moyenne de 13 sur 20.
Sans utiliser la formule de la moyenne, explique pourquoi c'est le cas.

c) Sabrina pense avoir une moyenne de 10 sur 20.
Sans utiliser la formule de la moyenne, explique pourquoi ce n'est pas le cas.

Fiche 2

Décrire l'impact d'un changement de données sur la moyenne

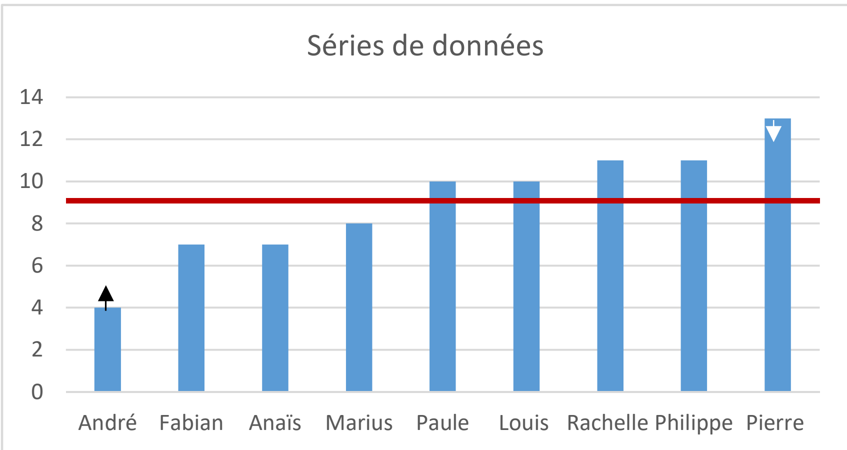
1. Lucie a obtenu les notes suivantes à ses contrôles de math : 16, 9 et 11.
   a) En utilisant l'appliquette GeoGebra, détermine la valeur de la moyenne de ses notes.
   b) Vérifie par calcul le résultat obtenu.
   c) Pour chacune des situations ci-dessous, entoure la proposition correcte.

Situation	A	B	C
Si elle avait eu moins de 16 à sa première note, sa moyenne …	aurait été supérieure à 12	aurait été inférieure à 12	aurait été égale à 12
Si elle avait eu 3 points de plus à l'une de ses notes, sa moyenne …	aurait augmenté de 3 points	aurait augmenté de 2 points	aurait augmenté de 1 point
Si elle avait eu 3 points de plus à chaque note, sa moyenne …	aurait augmenté de 3 points	aurait augmenté de 6 points	aurait augmenté de 9 points

   d) Vérifie les choix posés à la question précédente en utilisant l'appliquette.
   e) Explique, pour chacun des cas, la variation de la moyenne lorsqu'une donnée est modifiée.
   f) Si Lucie avait eu 6 points de moins à l'une de ses notes, comment aurait varié sa moyenne ?
   g) Si elle avait eu 1 point de moins à l'une de ses notes, comment aurait varié sa moyenne ?

Appliquette GeoGebra : https://www.geogebra.org/m/yt9ydyrh

2. Voici une série de données dont la moyenne vaut 9. Que devient la moyenne si on rapproche les deux valeurs extrêmes de la série d'une unité en direction de la moyenne ?



Fiche 3

Déterminer une donnée manquante dont la moyenne est donnée

1. Une série statistique comporte quatre valeurs. Sa moyenne vaut 4.
   Si trois de ces valeurs sont 3, 3 et 6, quelle est la valeur de la 4^{e} ?

2. Une série statistique comporte quatre valeurs. Sa moyenne vaut 4.
   Si trois de ces valeurs sont 1,2 et 5, quelle est la valeur de la 4^{e} ?

3. On a demandé à 5 élèves d'effectuer une collecte de fonds. S'ils reccueillent en moyenne 25€ chacun, ils gagnent des entrées pour assister à un match de basket. Lundi, 4 des 5 élèves se retrouvent et constatent qu'ils ont récolté : 29€, 21€, 31€ et 13€. Quel est le montant minimal que le dernier élève doit avoir reccueilli si le groupe veut gagner les entrées pour assister au match ?

4. Les dernières notes d'un élève dans une matière sont : 10, 13, 14, 9 et 8.
   Il espère remonter sa moyenne grâce au prochain devoir. Quelle note doit-il obtenir pour avoir au moins 11 de moyenne ce trimestre ?

5. Voici les notes sur 20 obtenues par Ludivine en mathématique.

Ludivine	13	16,5	17	13,5

Elodie a aussi été évaluée à 4 reprises. Elle a obtenu 4 fois la même note.
Elodie et Ludivine ont la même moyenne.
Quelles sont les notes d'Elodie en mathématiques ? Justifie.

Fiche élèves

Fiche 4

Exercices complémentaires

1. La moyenne des âges de Samia et d'Arthur est de 18 ans.
   La moyenne des âges de Samia et de Nicolas est de 15,5 ans.
   Combien d'années Arthur a-t-il de plus que Nicolas ?

2. Le matin du dernier jour des vacances, Julie se rend compte que si elle dépense encore 20 €, elle aura dépensé en moyenne 30 € par jour.
   Si elle dépense 70 €, cette moyenne sera de 40 €. Combien de jours ont duré les vacances de Julie ?

3. Dans une usine, parmi les ouvriers, on compte 46 femmes et 75 hommes. Le salaire moyen mensuel d'un ouvrier est de 1 982 €. Sachant que le salaire moyen d'un homme ouvrier est de 2 040 €, déterminer le salaire moyen des femmes ouvrières dans cette usine, arrondi à l'euro près.

ACTIVITÉ 2 : EXPLOITER LE MODÈLE DE LA BASCULE POUR COMPRENDRE LA MOYENNE

Cette activité développe une compréhension de la moyenne en tant que point d'équilibre d'une série de données. Cette compréhension est particulièrement aidante pour estimer une moyenne par exemple ou pour vérifier sa plausibilité.

L'introduction passe par une réflexion sur la plausibilité d'une moyenne calculée au départ d'un ensemble de données. L'exercice suivant, issu de l'évaluation non certificative peut servir d'amorce.

Voici le nombre de buts marqués par deux équipes lors de six matchs de handball.

Équipe 1	18	26	28	28	29	33
Équipe 2	27	24	32	23	28	25

a) Est-il possible que la moyenne de l'équipe 1 soit de 27 alors qu'elle n'a obtenu ce score à aucun match ? Explique ta réponse.

b) Un des joueurs de l'équipe 2 estime qu'en moyenne, son équipe a marqué environ 32 buts. Explique pourquoi, **sans calculer**, on peut dire que cette estimation est fausse.

	G	TT	TQ
Question a	59%	61%	45%
Question b	61%	53%	33%

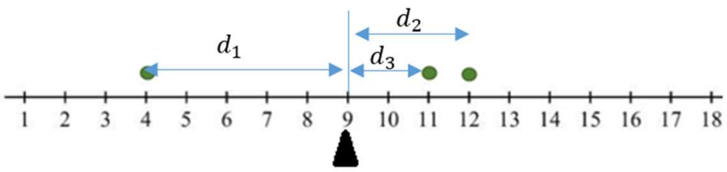
Après une discussion sur ces deux questions et une première estimation de la moyenne des deux équipes, la bascule est introduite : elle va aider les élèves à se construire une image de ce qu'est la moyenne. Cela leur permettra par exemple de vérifier qu'un résultat calculé par un logiciel est plausible.

L'activité se décline ensuite en deux parties.

Dans la première partie (fiches 5 et 6), les élèves explorent une bascule présentée sous une forme virtuelle, accessible à l'adresse suivante :

https://phet.colorado.edu/sims/html/balancing-act/latest/balancing-act_fr.html

D'un point de vue statistique, l'objectif est de définir les données d'une série statistique dont la moyenne vaut 9 (le point d'appui de la bascule virtuelle se situant à 9). L'exploitation des productions des élèves amène à constater que, pour obtenir une moyenne de 9, il est nécessaire que la somme des écarts à la moyenne soit nulle. Lors de cette exploitation, il peut être intéressant de visualiser les écarts à la moyenne à l'aide d'un schéma comme celui-ci :



$$d_1 = d_2 + d_3$$

Dans la deuxième partie de l'activité (fiches 7 et 8), l'élève cherche d'abord la moyenne d'une série en s'appuyant sur la propriété des écarts à la moyenne soit pour estimer la position de la moyenne, soit pour rechercher une situation symétrique où le point d'équilibre est aisé à trouver. Le premier exercice de la fiche 7 amorce cette seconde démarche. Dans la fiche 8, il approfondit les liens qui existent entre le calcul d'une moyenne simple et pondérée.

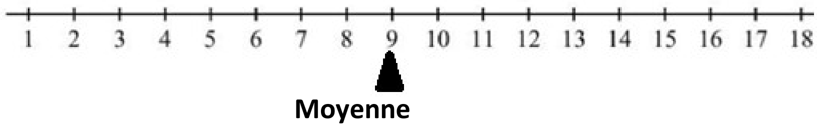
Au terme de l'activité, on synthétisera quelques propriétés de la moyenne :
- la moyenne d'une série de données correspond au point d'équilibre de cette série ;
- la moyenne est une valeur comprise entre le minimum et le maximum de la série de données ;
- quelle que soit la série, la somme des écarts à la moyenne s'élève à 0 et réciproquement.

On fera également le point sur les différentes formules pour calculer la moyenne, selon que l'on travaille avec des données brutes ou recensées.

Fiche 5 :

Générer une série de données dont la moyenne est donnée

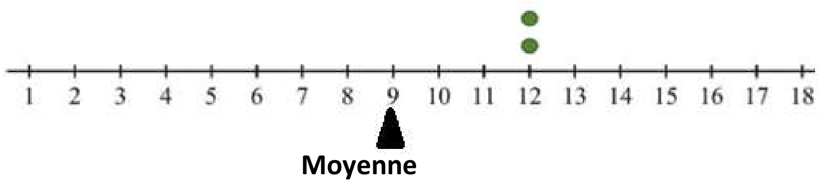
1) À l'aide de la bascule virtuelle, détermine un ensemble de 4 données dont la moyenne vaut 9.



Vérifie par calcul que la moyenne des quatre données vaut 9 :

____

2) Une série statistique a une moyenne de 9. Elle comporte 4 données. Deux données sont égales à 12. Utilise la bascule virtuelle pour déterminer les deux autres données.

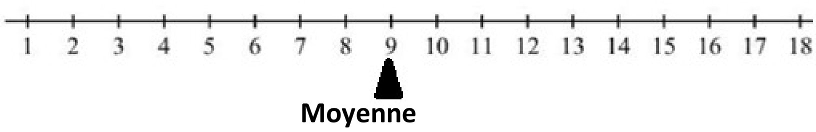


Vérifie par calcul que la moyenne des quatre données est égale à 9 :

____

3) Une série statistique a une moyenne de 9. Elle est composée de trois données : deux sont strictement supérieures à 9 et la dernière est inférieure à 9.

Utilise la bascule virtuelle pour déterminer les trois données de la série.



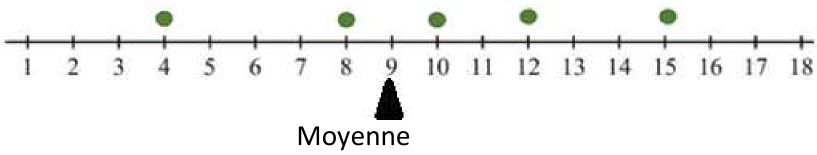
Vérifie par calcul que la moyenne des trois données est égale à 9.

____

Fiche 6

Quelques exercices complémentaires

1) La moyenne d'une série statistique vaut 9. Elle est composée de 6 données dont 5 sont connues : 4, 8, 10, 12 et 15.



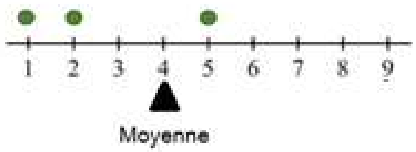
Détermine la donnée manquante : _______________

Vérifie que la moyenne de la série complète est égale à 9 en te basant sur les écarts à la moyenne.

____

____

2) Une série statistique comporte quatre valeurs. Trois de ces valeurs sont 1, 2 et 5. La moyenne de cette série est égale à 4.



Détermine la donnée manquante : __________________

Vérifie que la moyenne de la série complète est égale à 9 en te basant sur les écarts à la moyenne.

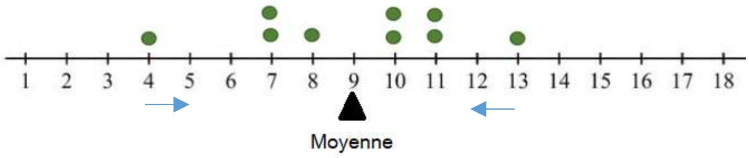
____

____

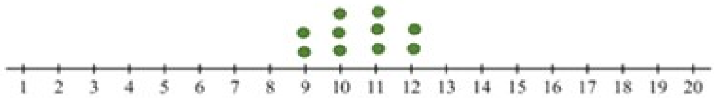
Fiche 7

Déterminer la moyenne d'un ensemble de données.

a) Voici une série de données dont la moyenne vaut 9. Que devient la moyenne si on rapproche les deux valeurs extrêmes de la série d'une unité en direction de la moyenne ?



b) Voici trois séries de données. Dans chaque cas, estime la moyenne et vérifie, par calcul, ta réponse.



Estimation de la moyenne : ________________________

Vérification par calcul : ____

Estimation de la moyenne : ………

Vérification par calcul :____

Estimation de la moyenne : ____________________

Vérification par calcul : ____

Fiche élèves

Fiche 8

Déterminer la moyenne d'un ensemble de données brutes ou recensées

La moyenne : des données brutes au tableau recensé

Deux dés à jouer ont été lancés 25 fois de suite.

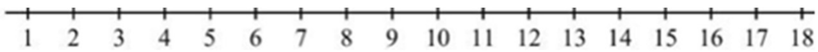
À chaque lancer, on a calculé la somme des points lus sur chaque dé.

Voici les sommes qui ont été calculées (données brutes) :

9	7	7	5	6
7	6	4	8	3
5	9	6	6	10
3	11	8	8	9
5	9	5	8	3

a) À partir de ce tableau de données, calcule la moyenne des sommes des points lus sur les dés. Écris le calcul réalisé.

b) Place les données sur le support ci-dessous. Détermine ensuite la position de la moyenne.



c) Voici un tableau recensé. Complète-le à partir de la liste des données brutes en t'aidant du support précédent.

Modalités (somme des 2 dés) x_i	Effectifs n_i
	N =

Interprète la troisième ligne du tableau.
Comment écris-tu le calcul de la moyenne au départ de ce tableau ?
Quelles sont les similitudes et/ou différences avec l'écriture du calcul réalisé au point a) ?

ACTIVITÉ 3 : LA MÉDIANE - LA COMPRENDRE ET LA CALCULER

L'activité 3 (fiches 9 à 11) vise à définir la médiane à partir d'observations réalisées à l'aide de l'appliquette GeoGebra.

Lorsque l'élève introduit les données, l'appliquette GeoGebra détermine la valeur de la médiane et en donne une représentation. À travers les différentes situations données, l'élève décrit les caractéristiques de la médiane pour ensuite en donner une définition.

L'élève manipule progressivement :

- un nombre impair de données différentes ;
- un nombre pair de données différentes ;
- un ensemble de données dont certaines sont égales.

On propose à l'enseignant de construire avec les élèves la définition de la médiane en s'appuyant sur les observations réalisées dans les trois premiers cas.

La quatrième situation est réalisée sans l'appliquette afin que l'élève prenne conscience de l'utilité d'ordonner les données.

On fera également le point sur les différentes méthodes pour déterminer la médiane, selon que l'on travaille avec des données brutes ou recensées.

Fiche élèves

Fiche 9 :

Caractériser et définir la médiane

1) Dans les trois cas ci-dessous :
    - encode dans l'appliquette GeoGebra les données reprises dans le tableau et écris la valeur de la médiane ;

      https://www.geogebra.org/m/ttprampf
    - observe comment se répartissent les données autour de la médiane ;
    - généralise au moyen d'une phrase les observations réalisées.

1er cas : Nombre impair de données différentes

Données	Médiane	Observations
3 ; 7 ; 10		
4 ; 12 ; 6		
2 ; 13 ; 14 ; 7 ; 9		
5 ; 6 ; 7 ; 8 ; 9		
3 ; 10 ; 15 ; 7 ; 1 ; 8 ; 11		

Généralisation :

2^{e} cas : Nombre pair de données différentes

Données	Médiane	Observations
3 ; 7 ; 10 ; 11		
4 ; 12 ; 6 ; 2		
2 ; 13 ; 14 ; 7 ; 9 ; 11		
5 ; 6 ; 7 ; 8 ; 9 ; 10		
3 ; 10 ; 15 ; 7 ; 1 ; 8 ; 11 ; 4		

Généralisation :

3^{e} cas : Situations particulières

Données	Médiane	Observations
3 ; 3 ; 3		
5 ; 5 ; 7 ; 10 ; 14		
4 ; 6 ; 6 ; 10 ; 12		
5 ; 9 ; 9 ; 11		
5 ; 5 ; 5 ; 8		
8 ; 8 ; 9 ; 12 ; 16 ; 16		

Généralisation :

2) Sans l'appliquette, détermine la médiane en t'appuyant sur la définition

Données	Médiane
1 ; 4 ; 5	
8 ; 2 ; 3	
2 ; 7 ; 6 ; 4 ; 10	
3 ; 5 ; 7 ; 9 ; 11	
7 ; 2 ; 4 ; 8	
4 ; 8 ; 9 ; 12	
1 ; 4 ; 5 ; 8 ; 9 ; 14	
9 ; 2 ; 5 ; 11 ; 18	

Fiche élèves

Fiche 10 :

S'exercer à déterminer la médiane

Situation :

Cinq amis se cotisent pour offrir un cadeau, ils donnent chacun respectivement 26 € ; 20 € ; 24 € ; 22 € et 28 €. Détermine la médiane.

À partir de cette situation, complète le tableau.

Contexte	Ensemble de données	Médiane	Observations
Un des amis ne donne plus 20 € mais 10 €.			
Ils décident tous de donner 2 € supplémentaires.			
Le plus petit montant diminue de 2 € et le plus grand montant augmente de 2 €.			
Deux autres amis se rajoutent ; ils mettent 10 € et 11 €			
L'ami qui avait donné 28 € décide de se retirer de la cagnotte.			
L'ami qui avait donné 20 € décide de se retirer et l'ami qui avait donné 26 € décide de donner 2 € de moins.			
L'ami qui avait mis 24 € décide de mettre 35 €.			

Fiche élèves

Fiche 11

Déterminer la médiane d'un ensemble de données brutes ou recensées

La médiane : des données brutes au tableau recensé

Deux dés à jouer ont été lancés 25 fois de suite.

À chaque lancer, on a calculé la somme des points lus sur chaque dé.

Voici les sommes qui ont été calculées :

9	7	7	5	6
7	6	4	8	3
5	9	6	6	10
3	dia	8	8	9
5	9	5	8	3

a) À partir de ce tableau de données, détermine la médiane.
Explicite ta démarche.

b) Complète le tableau recensé de cette série statistique en ordre croissant.

Modalités (somme des 2 dés) x_i	Effectifs n_i	Effectifs cumulés
	N =	

Interprète la troisème ligne de ce tableau.
Comment détermines-tu la médiane à partir de ce tableau ?

https://www.geogebra.org/m/qcfsjvpt

ACTIVITÉ 4 : MOYENNE ET MÉDIANE, QUE CHOISIR ?

L'activité 4 (fiche 12) vise à amener les élèves à prendre conscience des avantages respectifs de la médiane et de la moyenne pour décrire la tendance centrale d'une série statistique. Certaines situations impliquent des variables qualitatives où seul le calcul de la médiane est pertinent. D'autres situations amènent les élèves à argumenter sur le choix de la moyenne ou de la médiane lorsqu'il y a des valeurs atypiques.

Fiche élèves

Fiche 12

Argumenter à partir d'une moyenne et d'une médiane

1) Dans une école, l'évaluation ne se base pas sur des points mais sur des appréciations :

TI (très insuffisant) ; I (insuffisant) ; F (faible) ; S (satisfaisant) ; B (bien) ; TB (très bien) ; R (remarquable).

Voici les résultats des élèves pour un examen oral de mathématiques :

Adrien	S	Katherine	S
Amandine	B	Loïc	B
Amaury	S	Luigi	I
Arthur	F	Maëlle	S
Bechara	F	Maria-Isabel	F
Celia	B	Mariam	S
Hélène	B	Nawal	B
Hugues	TB	Nesrine	B
Jules	TI	Pierre	TB
Julien	I	Valentino	R
Karim	S	Victor	S

Est-il possible de déterminer la moyenne des notes obtenues par les élèves de la classe ?

Si oui, donne cette moyenne. Sinon, justifie pourquoi ce n'est pas possible.

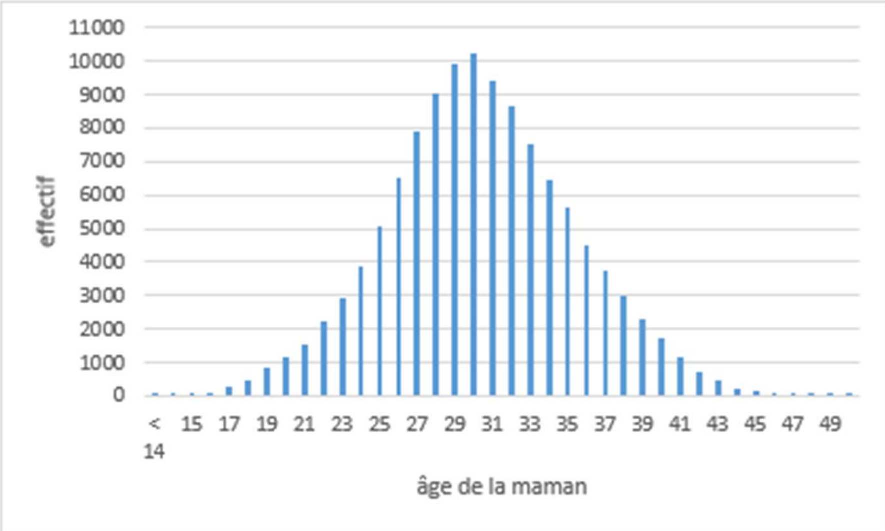
Est-il possible de déterminer la médiane des notes obtenues par les élèves de la classe ?

Si oui, donne cette médiane. Sinon, justifie pourquoi ce n'est pas possible.

2) Six lecteurs achètent le dernier livre de leur auteur favori.
Ils paient respectivement 45 €, 8 €, 9 €, 7 €, 8 €, 10 €.
Est-il plus intéressant de prendre la moyenne ou la médiane pour caractériser la série statistique ?
Justifie ta réponse.

3) Commente chacune des situations présentées par les deux diagrammes ci-dessous en te basant sur les valeurs de la moyenne et de la médiane.

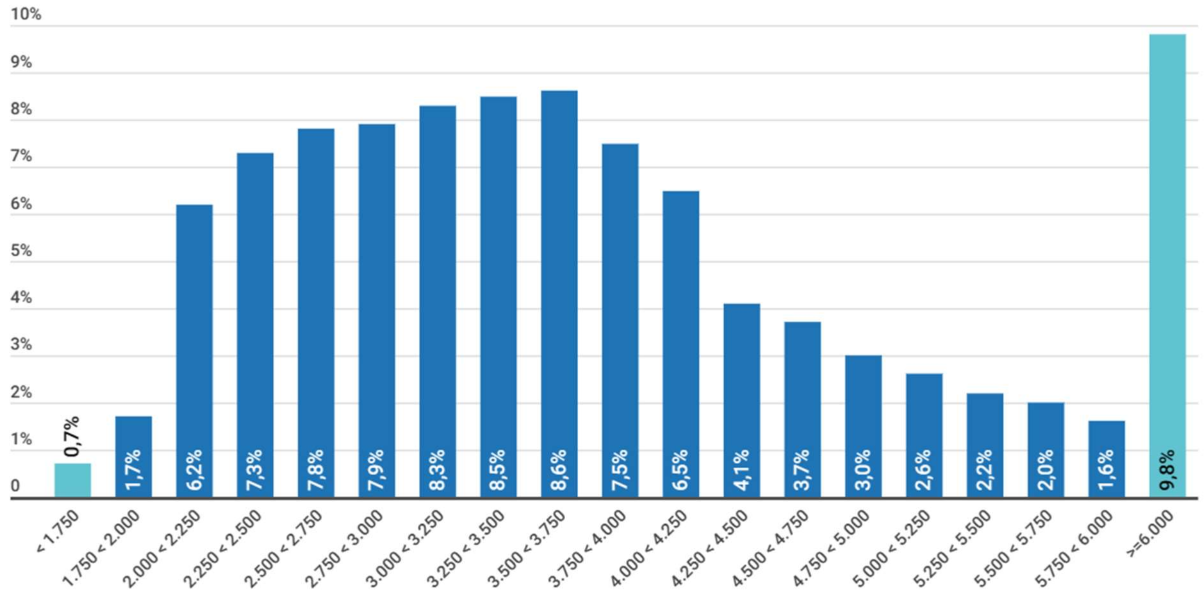
Le diagramme 1 présente la répartition des naissances en Belgique en 2018 selon l'âge de la mère.



Une mère a en **moyenne** 30,4 ans à la naissance de son enfant.

L'âge **médian** d'une maman à la naissance de son enfant est de 30 ans.

Le diagramme 2 présente la répartition des salaires mensuels bruts moyens en Belgique pour l'année 2020.



En 2020, le salaire mensuel brut **moyen** d'un salarié employé à temps plein s'élevait à 3.832 euros.

En 2020, le salaire mensuel brut **médian** d'un salarié employé à temps plein s'élevait à 3.550 euros.

Source : https://statbel.fgov.be/fr/themes/emploi-formation/salaires-et-cout-de-la-main-doeuvre/salaires-mensuels-bruts-moyens#:~:text=1.,s'%C3%A9levait%20%C3%A0%203.832%20euros.

PARTIE 2 - DES LIMITES DES INDICATEURS DE TENDANCE CENTRALE À L'ÉCART-TYPE

1. LES CONSTATS ISSUS DE L'ÉPREUVE

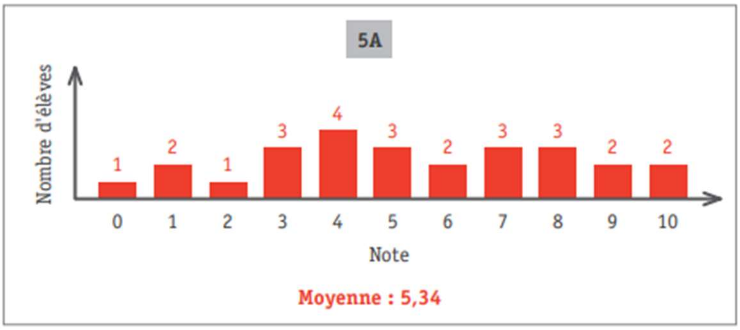
L'épreuve ayant été soumise aux élèves au début de l'année scolaire 2022-2023, peu de questions portent sur les indicateurs de dispersion.

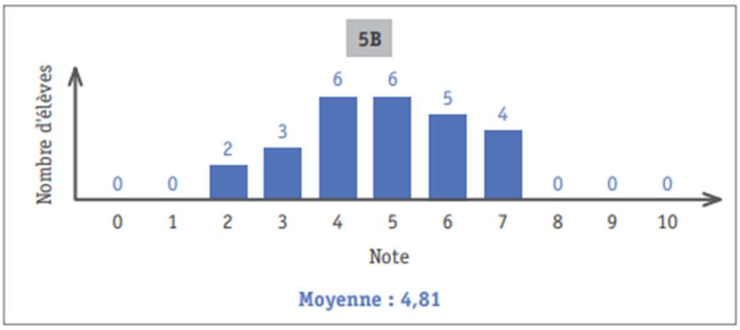
Toutefois, lorsque nous avons interrogé les enseignants de 5e secondaire de l'échantillon sur les thématiques à approfondir dans les pistes, ils ont confirmé le fait que les indicateurs de dispersion, et en particulier l'écart-type, sont des prérequis importants pour aborder les statistiques à deux variables dans l'enseignement de transition et de qualification. C'est la raison pour laquelle le groupe de travail a décidé de proposer une réflexion sur ce thème.

Dans cette section, nous nous attarderons brièvement sur une question de l'épreuve proposée aux élèves de transition :

QUESTION 13

Un test noté sur 10 points a été proposé en 5A et en 5B. Voici la répartition des points attribués aux élèves.





a) Explique pourquoi l'écart-type pour la 5A est plus grand que celui pour la 5B. ☐ 37

b) Détermine l'étendue pour la 5B : ______ ☐ 38

	G	TT
Item 37	39%	35%
Item 38	22%	16%

Ces deux items figurent parmi les plus complexes de l'épreuve.

Selon plusieurs chercheurs (Delmas & Liu, 2005), ce type de difficultés des élèves est lié à une compréhension superficielle de concepts tels quel le centre, la distribution et la dispersion des données. La dispersion des données autour de la moyenne n'est pas un critère qu'ils prennent en compte lorsqu'ils analysent des distributions. Or c'est une compétence importante pour donner sens à l'écart-type. Cette idée est développée dans la suite. En ce qui concerne l'item 37, nous constatons un taux élevé d'omission (17% en G et 20% en TT). Certains se sont exclusivement basés sur la comparaison des moyennes pour justifier leur choix, sans évoquer la dispersion des données autour des deux moyennes. Le concept d'écart-type est donc très peu maitrisé : cela pourrait s'expliquer par le contexte sanitaire. En ce qui concerne l'étendue, les élèves ont tendance à l'appréhender au travers de la hauteur des bâtonnets, lorsque les élèves travaillent sur des diagrammes en bâtonnets, comme c'est le cas dans cet exercice : pour l'item 38, plusieurs élèves ont considéré que l'étendue pour la 5B était égale à 4 car le bâtonnet le plus petit a une hauteur de 2 et le bâtonnet le plus grand a une hauteur de 6. Nous avons également observé que de nombreux élèves ont tout simplement omis de répondre à cet item (26% d'omission en G et 31% d'omission en TT). Il est donc difficile d'avoir accès à leurs démarches même tâtonnantes. Quelques élèves enfin ont identifié le maximum et le minimum de la série de données (7 et 2), sans penser à calculer leur différence.

En référence à ces résultats, cette seconde partie des pistes didactiques propose des activités visant à dépasser les indices de tendance centrale, pour explorer plus finement la notion d'écart-type.

2. INTENTIONS ET COMMENTAIRES

Afin d'amener les élèves à mieux comprendre la notion d'écart-type, tant dans ses aspects conceptuels que calculatoires, trois activités sont proposées.

La **première activité** introduit l'importance de s'attacher à la dispersion des données. Elle montre les limites des indicateurs de tendance centrale, en particulier de la moyenne et de la médiane, pour décrire une série statistique. Cette attention sur la dispersion des données est un prérequis important pour percevoir la nécessité de développer des indicateurs de dispersion, et en particulier la notion d'écart-type.

La **deuxième activité** décortique pas à pas la formule de l'écart-type en amenant les élèves à comprendre comment la notion d'écart à la moyenne est prise en compte dans le calcul de l'écart-type et évolue vers le calcul de la variance (dont l'unité ne correspond pas à celle de la série de données), puis de l'écart-type.

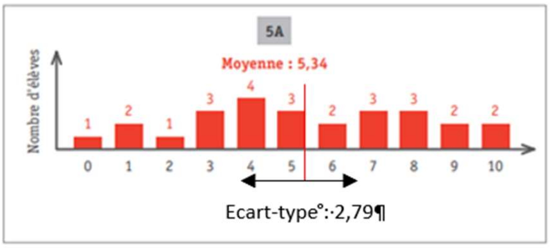
La **troisième activité,** enfin, afine le travail sur l'écart-type en lien avec la densité relative des données autour de la moyenne.

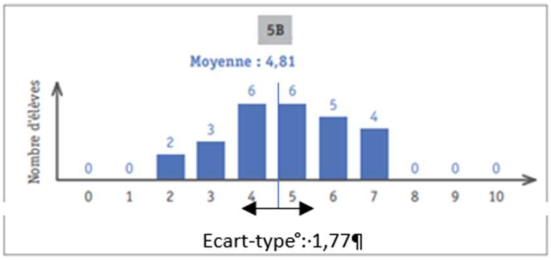
L'écart-type

Une compréhension approfondie de l'écart-type passe par la mise en perspective de trois concepts (Delmas & Liu, 2005).

- Le premier concept est celui de l'analyse de la manière dont les données se ditribuent : sont-elles homogènes ou plutôt hétérogènes ? On s'intéresse donc aux fréquences et aux fréquences cumulées d'une série statistique.
- Un second concept fondamental est celui de la moyenne appréhendée non pas au travers de sa formule, mais plutôt envisagée comme le point d'équilibre des données (comme développé dans la première partie de ces pistes). Cette image de la moyenne permet de donner du sens à ce concept en tant qu'indicateur de tendance centrale.
- Un troisième concept est important : la dispersion par rapport à la moyenne. En effet, c'est bien une quantification de cette dispersion qui constitue la raison d'être de l'écart-type et de la variance.

C'est donc en combinant une analyse de la distribution (en s'intéressant aux fréquences) et de la dispersion (en tant que distance par rapport à la moyenne), que l'on obtient une compréhension approfondie de l'écart-type, envisagé comme la densité relative des données autour de la moyenne. Un élève qui possède cette compréhension peut anticiper la manière dont les valeurs possibles d'une variable et leurs fréquences respectives influencent l'écart-type. Cela peut être visualisé graphiquement :





À terme, ce travail sur l'écart-type débouche sur l'inégalité de Tchebychev qui constitue l'un des outils d'interprétation de l'écart-type (Henry & Miewis, 2016, p.38) :

> Pour toute série statistique de moyenne $\bar{x}$ et d'écart-type σ, la proportion de données appartenant à l'intervalle $[\bar{x} - \sigma, \bar{x} + \sigma]$ est supérieure ou égale à $1 - \frac{1}{t^2}$, avec $t > 0$

Ce théorème permet de déterminer la proportion de données situées dans un intervalle autour de la moyenne, en se basant uniquement sur l'effectif, la moyenne et l'écart-type. Henry et Miewis (2016) préconisent de faire varier les différents éléments de l'énoncé (moyenne, écart-type et/ou effectif) en vue de prendre pleinement conscience de l'influence de ces paramètres sur l'intervalle autour de la moyenne.

Une entreprise est chargée de produire 500 pièces. La commande stipule que ces pièces doivent peser 2 kg, mais une tolérance de 100 grammes en plus ou en moins est acceptée. En moyenne, les pièces pèsent 2kg et l'écart-type s'élève à 10 grammes. Combien de pièces remplissent les conditions de la commande ?

Face au problème ci-dessus, les élèves peuvent déduire de l'inégalité de Tchebytchev que 99% au moins des pieds de table remplissent les conditions.

En effet, l'intervalle étant connu, on peut déduire la valeur de t

$$\bar{x} - ts = 1900$$
$$\Leftrightarrow$$
$$2000 - t \cdot 10 = 1900$$
$$\Leftrightarrow$$
$$t = 10$$

La proportion d'observations est donc strictement supérieure à $1 - \frac{1}{t^2}$, soit 99%

En faisant varier les données de l'énoncé (écart-type de 15 grammes, tolérance réduite à 50 g autour de la moyenne), les élèves peuvent prendre conscience de ces changements sur le nombre ou la proportion de pièces répondant aux critères de tolérance définis.

ACTIVITÉ 5 : LIMITES DES INDICES DE POSITION POUR ANALYSER UNE DISTRIBUTION

Cette activité vise à faire réfléchir les élèves sur la variabilité de séries statistiques envisagées dans des contextes différents: ils s'interrogent sur la pertinence de la moyenne ou de la médiane pour soutenir une décision.

L'activité se structure en deux parties.

Dans la première partie (fiches 1 et 2), les élèves découvrent les situations à analyser. Ils préparent leurs arguments en faveur ou en défaveur de l'indicateur proposé pour prendre la décision.

Dans la deuxième partie, le débat contradictoire est mené par l'enseignant : les élèves sont alors amenés à formuler les limites des indicateurs proposés pour soutenir la décision à prendre et à envisager d'autres types d'analyse.

L'activité se termine par une mise au point sur les différents indicateurs statistiques, et la nécessité de s'intéresser à la dispersion des données pour analyser une série statistique.

Première partie : Préparation du débat contradictoire (fiches 1 et 2)

Les élèves sont répartis en groupes de quatre. Ils découvrent les différentes situations de la fiche 1 et l'enseignant leur attribue l'analyse d'une d'entre elles. Ensuite, chaque groupe découvre les consignes pour le travail (voir fiche 2) : il se scinde en deux duos : le premier duo doit défendre la proposition formulée dans la situation et le second duo doit trouver des arguments contre celle-ci.

Brève analyse des situations :

1. Un responsable d'un mouvement de jeunesse prépare une sortie au bowling avec 30 personnes. Étant donné la taille importante du groupe de joueurs, il doit réserver des paires de chaussures à enfiler avant de commencer les parties. Pour ce faire, il relève la pointure de chaque jeune et décide de réserver 30 paires de chaussures dont la pointure correspond à la moyenne calculée.

   <u>Analyse</u> : Cette situation peut faire appel à la notion d'effectif. La moyenne ne sera pertinente qu'au cas où tous les élèves ont la même taille de chaussures, ce qui est peu plausible.

2. Le 27 septembre 2022, Le Soir[4] publie un article dont une information principale est le salaire médian de belges : « *En Belgique, le salaire médian était de 3550 euros bruts par mois en 2020*. » Un patron d'entreprise engage un nouvel employé et décide de le rémunérer avec ce salaire mensuel de 3550 euros brut.

   <u>Analyse</u> : La notion de médiane se justifie dans la mesure où la distribution des salaires bruts présente souvent des salaires atypiques. Toutefois, le choix du salaire à accorder à un nouvel employé dépend de bien d'autres facteurs que ceux purement statistiques (l'ancienneté de la personne, ses diplômes, les possibilités de l'entreprise, …)

3. Pour le 22 mars 2023, l'Institut royal météorologique de Belgique (IRM) annonce une température moyenne de cette journée de 14°. Nour, élève de 4e secondaire, estime cette température suffisamment agréable et décide de partir à l'école sans manteau.

[4] https://www.lesoir.be/467835/article/2022-09-27/quel-est-le-salaire-median-des-belges-infographies?referer=%2Farchives%2Frecherche%3Fdatefilter%3Dlastyear%26sort%3Ddate%2Bdesc%26start%3D30%26word%3Dsalaire%2Bm%25C3%25A9dian, consulté le 25 janvier 2023.

<u>Analyse</u> : Cette situation peut faire appel à la notion d'étendue, de minimum et de maximum. Par ailleurs d'autres données, telles que le risque de précipitations, pourraient être utiles pour prendre la meilleure décision.

4. Un professeur a corrigé les interrogations des 28 élèves de sa classe. Il a retenu que la moyenne des notes vaut 12,4/20, mais on lui a volé sa mallette dans le coffre de sa voiture. Il ne peut donc pas donner à chaque élève son propre résultat et décide d'attribuer la note moyenne à tout le monde.

   <u>Analyse</u> : Cette situation fait appel à la dispersion des notes. La moyenne pourrait être pertinente si tous les élèves ont en réalité obtenu la même note ou, en pratique, ne fût-ce qu'approximativement la même note.

5. Le gouvernement d'un pays désire soutenir les familles avec une prime de 200€ par enfant. Étant donné que le nombre moyen d'enfants par femme est de 1,78, le gouvernement décide de donner une prime de 356€ à chaque famille.

   <u>Analyse</u> : Cette situation invite à une réflexion sur le caractère équitable de la proposition ainsi que sur le sens de la moyenne dans ce contexte. Une analyse de la dispersion des résultats serait utile pour décider d'une mesure plus équitable visant à répartir différemment le montant total à octroyer.

6. Douze amis passent une soirée au restaurant autour d'un bon repas. Lorsque l'addition est présentée, le montant total s'élève à 624€. L'organisateur du rendez-vous décide de demander 52€ à chaque personne afin de payer cette addition.

   <u>Analyse</u> : Si tout le monde a pris le même repas, cela ne pose aucun problème. C'est à nouveau la dispersion qui devrait être analysée pour prendre la décision la plus juste.

Deuxième partie : Mise en commun

En fonction du temps disponible pour réaliser la mise en commun, on organisera un débat collectif autour de l'une ou de l'autre situation. Les élèves qui ne participent pas activement au débat pourront, au terme de celui-ci, discuter des différents arguments avancés. Il ne parait en effet pas nécessaire de passer en revue toutes les situations, car plusieurs d'entre elles abordent des thématiques proches.

L'activité se termine par une synthèse visant à mettre en évidence les limites de la moyenne et de la médiane pour décrire une distribution statistique.

Fiche 1 :

Choisir le bon indicateur de tendance –
Quand les indices de position ne suffisent pas

Voici une série de situations, impliquant une prise de décision (indiquée en gras).

Analyse l'une d'entre elles, en te référant aux consignes de la fiche 2.

1. Un responsable d'un mouvement de jeunesse prépare une sortie au bowling avec 30 personnes. Étant donné la taille importante du groupe de joueurs, il doit réserver des paires de chaussures à enfiler avant de commencer les parties. Pour ce faire, il relève la pointure de chaque jeune et **décide de réserver 30 paires de chaussures dont la pointure correspond à la moyenne calculée**.
2. Le 27 septembre 2022, Le Soir[5] publie un article dont une information principale est le salaire médian de belges : *En Belgique, le salaire médian était de 3550 euros bruts par mois en 2020*. Un patron d'entreprise engage un nouvel employé et **décide de le rémunérer avec ce salaire mensuel de 3550 euros brut**.
3. Pour le 22 mars 2023, l'Institut royal météorologique de Belgique (IRM) annonce une température moyenne de cette journée de 14°. Nour, élève de 4e année, estime cette température suffisamment agréable et **décide de partir à l'école sans manteau**.
4. Un professeur a corrigé les interrogations des 28 élèves de sa classe. Il a retenu que la moyenne des notes vaut 12,4/20, mais on lui a volé sa mallette dans le coffre de sa voiture. Il ne peut donc pas donner à chaque élève son propre résultat et **décide d'attribuer la note moyenne à tout le monde**.
5. Le gouvernement d'un pays désire soutenir les familles avec une prime de 200€ par enfant. Étant donné que le nombre moyen d'enfants par femme s'élève à 1,78, le gouvernement **décide de donner une prime de 356€ à chaque famille**.
6. Douze amis passent une soirée au restaurant autour d'un bon repas. Lorsque l'addition est présentée, le montant total s'élève à 624€. L'organisateur du rendez-vous **décide de demander 52€ à chaque personne afin de payer cette addition**.

[5] https://www.lesoir.be/467835/article/2022-09-27/quel-est-le-salaire-median-des-belges-infographies?referer=%2Farchives%2Frecherche%3Fdatefilter%3Dlastyear%26sort%3Ddate%2Bdesc%26start%3D30%26word%3Dsalaire%2Bm%25C3%25A9dian, consulté le 25 janvier 2023.

Fiche 2 :

Consignes pour préparer le débat contradictoire

1) Cochez ci-dessous le n° de la situation qui vous a été attribuée :

   Mon groupe doit travailler sur la situation n°

1	2	3	4	5	6

2) Au sein de votre groupe, formez deux sous-groupes.
   L'un devra **défendre** la proposition donnée, l'autre devra trouver des **arguments contre** celle-ci. Ce second sous-groupe devra également faire une autre proposition permettant de répondre à la question posée.

   Indiquez ci-dessous le nom des élèves qui doivent trouver des arguments pour et contre.

POUR	CONTRE
Nom de l'élève 1 : …………………………	Nom de l'élève 1 : …………………………
Nom de l'élève 2 …………………………	Nom de l'élève 2 …………………………

   Pour cela, vous devrez utiliser le vocabulaire statistique parmi ces propositions :

   donnée, individu, population, échantillon, variable, valeur, observation, étendue, effectif, fréquence, pourcentage, moyenne, médiane, mode, distribution, dispersion, représentativité, certain, impossible, probabilité, cas particulier, généralité, contre-exemple, équitable

3) Sur une feuille de brouillon, préparez vos arguments.

Temps avant mise en commun : ……. Min

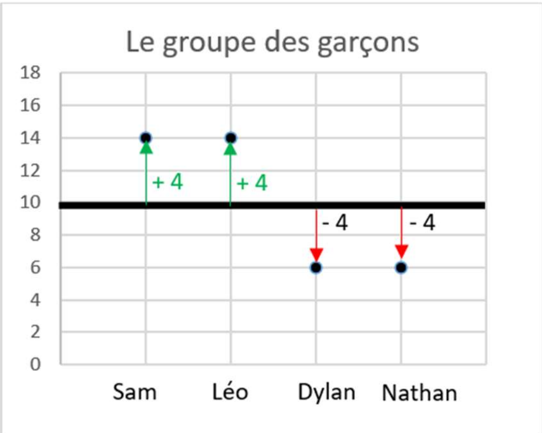
ACTIVITÉ 6 : COMPRENDRE LA FORMULE DE L'ÉCART-TYPE

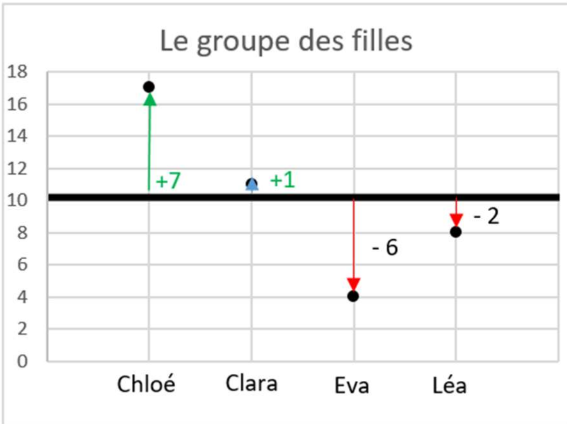
Cette activité vise à approcher pas à pas la notion d'écart-type. Elle se structure en trois parties.

La première partie (fiche 3) amène les élèves à comparer les résultats de deux classes en calculant d'abord les indices de tendance centrale (qui sont tous identiques) puis en analysant des graphiques présentant la distribution des notes dans les deux classes. Ils prennent alors conscience que les deux distributions sont contrastées : dans l'une, les résultats sont plus homogènes que dans l'autre. L'écart-type est alors présenté comme l'indicateur qui va permettre de quantifier ce caractère plus ou moins homogène des données autour de la moyenne.

La deuxième partie (fiche 4) travaille ensuite la formule de l'écart-type au départ de quelques données. L'idée est de faire réfléchir les élèves sur la manière de quantifier les écarts à la moyenne de deux distributions, dont l'une est moins homogène que l'autre. Ces deux distributions sont présentées dans un contexte de prix, impliquant donc des euros. C'est important pour faire réfléchir les élèves au sens de la variance dans ce contexte (des euros au carré n'ont en effet pas de sens).

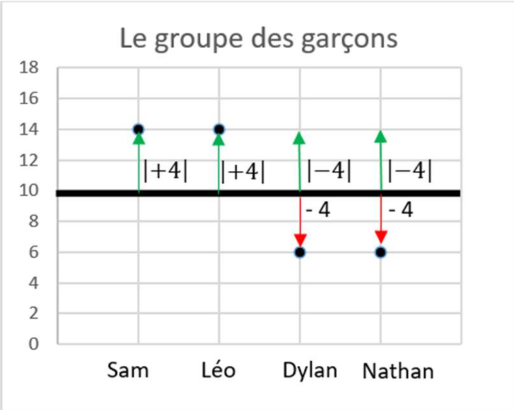
- Les élèves calculent d'abord la somme des écarts à la moyenne et constatent que, dans les deux cas, le résultat est égal à 0. Ils revoient ainsi une propriété importante de la moyenne déjà abordée dans la première partie de ces pistes.

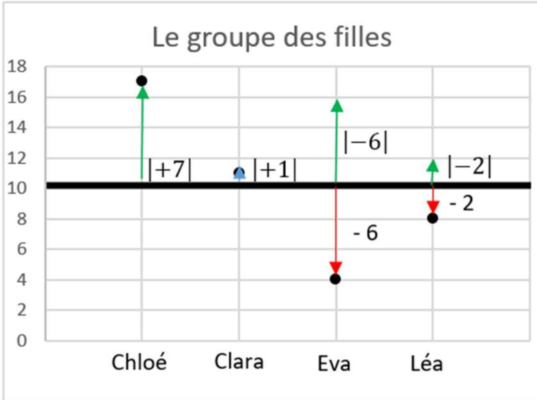




- Ils réfléchissent aux possibilités de quantifier les écarts à la moyenne, pour tenir compte à la fois des écarts positifs et négatifs.
  Deux possibilités pourraient être avancées :

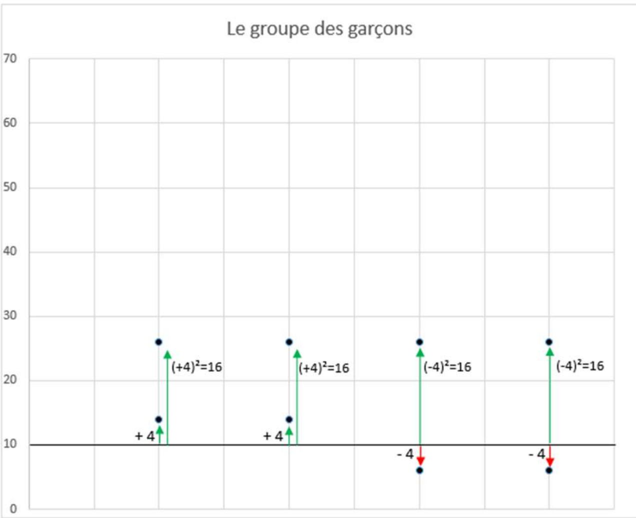
  - calculer la moyenne des écarts en se basant sur les valeurs absolues :

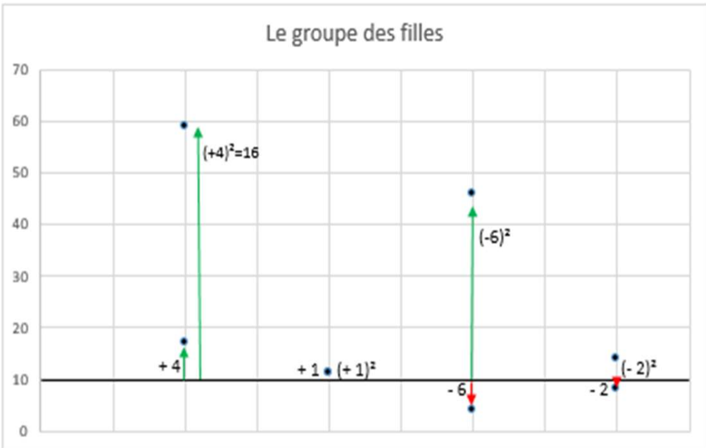




Les élèves calculent l'écart moyen. C'est une mesure utilisée notamment dans le domaine des finances. D'un point de vue mathématique, la fonction « valeur absolue » est plus difficile à manipuler ; c'est une des raisons pour lesquelles une autre alternative a été trouvée pour éviter que les valeurs négatives ne compensent les valeurs positives.

  - L'autre possibilité est d'élever toutes les valeurs au carré, et de calculer ainsi la variance.





En comparant les graphiques obtenus, selon qu'on prenne en considération les écarts en valeur absolue ou le carré des écarts, on peut montrer que, dans le cadre du calcul de la variance, on donne plus d'importance aux valeurs particulièrement éloignées de la moyenne qu'à celles qui lui sont proches.

En réfléchissant à l'unité obtenue lorsqu'on travaille avec le carré des écarts, les élèves peuvent prendre conscience du manque de sens de l'unité de la variance : des euros au carré n'existent pas. Il convient dès lors de prendre la racine carrée de la variance pour avoir une donnée exprimée dans la même unité que les données de départ.

La troisième partie (fiche 5) invite enfin les élèves à comparer le caractère plus ou moins homogène de trois distributions ayant une même moyenne. Cette troisième partie réinvestit, dans un contexte de grands nombres, les différentes étapes nécessaires pour calculer les écart-types de trois distributions. Avant de procéder aux calculs à l'aide d'un tableur, les élèves sont amenés à anticiper la position relative des trois écart-types.

La suite de la fiche présente les données permettant aux élèves de visualiser les étapes nécessaires pour déterminer la variance et l'écart-type sur un tableur. L'enseignant pourrait projeter au tableau les deux premières colonnes du tableau et ensuite :

- pour la première distribution, construire avec eux l'ensemble du tableau sur Excel en les faisant réfléchir sur la raison pour laquelle les écarts sont élevés au carré (influence sur les unités et sur la dispersion des écarts) ;
- pour la deuxième distribution, réaliser l'exercice sur Excel **et** déterminer la moyenne et l'écart-type à l'aide de la machine à calculer ;
- pour la troisième distribution, déterminer la moyenne et l'écart-type uniquement à l'aide de la machine à calculer.

Fiche 3 :

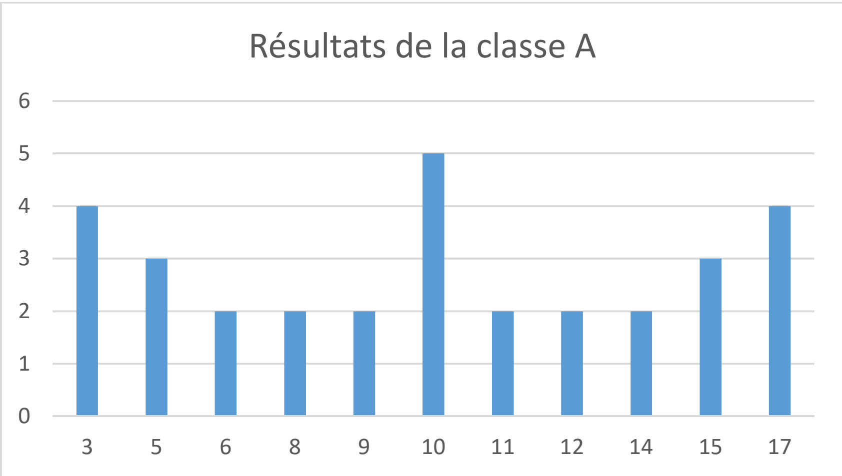
Décrire deux ensembles de données

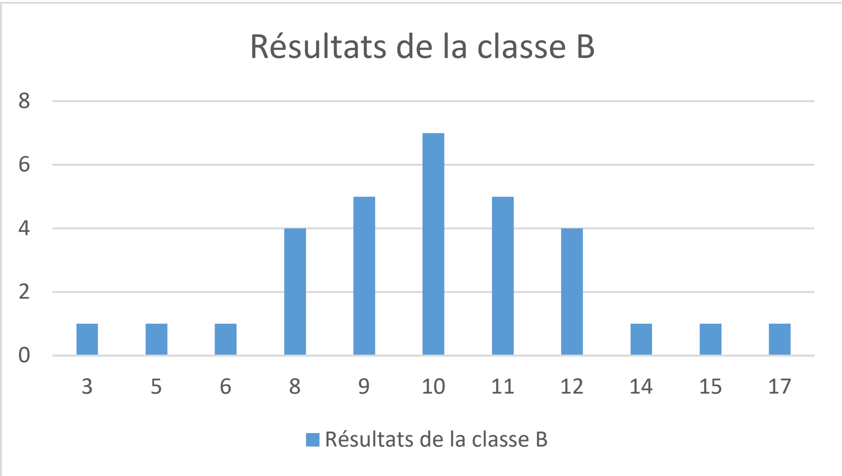
Voici les résultats à un test de mathématiques de deux classes.

Note	**Effectif (classe A)**	**Effectif (classe B)**
3	4	1
5	3	1
6	2	1
8	2	4
9	2	5
10	5	7
11	2	5
12	2	4
14	2	1
15	3	1
17	4	1
TOTAL	**31**	**31**

1) Compare les résultats de ces deux classes en calculant la moyenne. Est-il possible de distinguer ces deux classes à l'aide de cet indicateur ?

2) Observe à présent les graphiques des effectifs de ces deux classes. En quoi les résultats sont-ils différents ?





Fiche 4 :

Analyser les écarts à la moyenne d'une distribution

Deux groupes d'amis organisent une « cacahuète » : ils décident de s'offrir un cadeau mutuellement, d'une valeur moyenne de 10 euros.

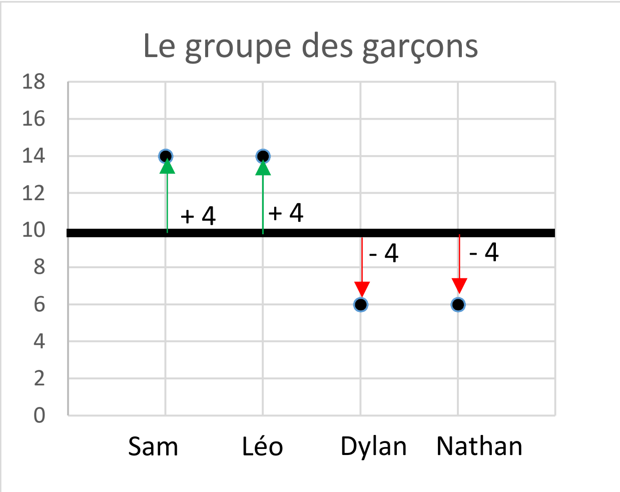
Voici le prix des cadeaux pour chaque groupe d'amis :

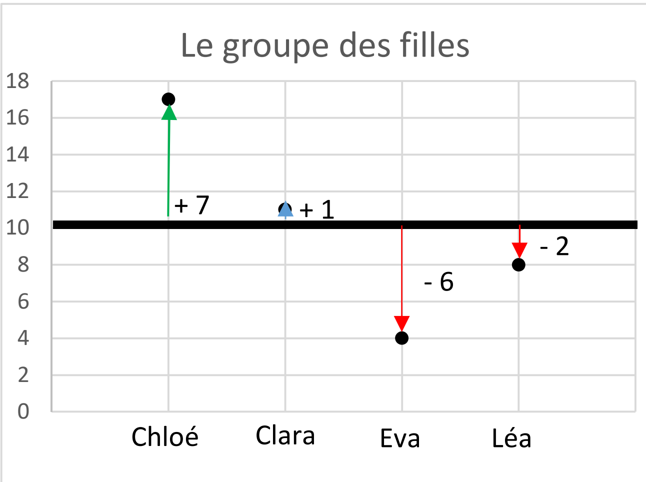
Le groupe des garçons	
Sam	14 €
Léo	14 €
Dylan	6 €
Nathan	6 €
Prix moyen	10 €

Le groupe des filles	
Chloé	17 €
Clara	11 €
Éva	4 €
Léa	8 €
Prix moyen	10 €

Les écarts entre chaque prix et la moyenne des prix sont différents. L'écart-type est un indice statistique qui va permettre de quantifier la moyenne de ces écarts à la moyenne.

Le graphique suivant permet de visualiser les écarts à la moyenne, dans chaque groupe.





L'écart-type est un paramètre qui vise à déterminer la moyenne des écarts à la moyenne.

Plusieurs démarches pourraient être envisagées pour déterminer cette moyenne des écarts.

1) Calcule la moyenne de ces écarts, en distinguant les écarts positifs (la donnée est supérieure à la moyenne) et négatifs (la donnée est inférieure à la moyenne).

    a. Pour les garçons :

    b. Pour les filles :

2) Une solution pour éviter que les écarts positifs compensent les écarts négatifs est de travailler avec les valeurs absolues. Effectue les calculs en travaillant cette fois avec les valeurs absolues.

    a. Pour les garçons :

    b. Pour les filles :

3) Il existe encore une autre façon d'éviter de travailler avec des valeurs négatives : élever chaque écart au carré.
   Effectue les calculs en travaillant cette fois avec le carré de chaque écart.

    a. Pour les garçons :

    b. Pour les filles :

Laquelle des trois démarches effectuées te permet de rendre le mieux compte de la dispersion des résultats par rapport à la moyenne ? …………..

La variance a été définie au départ de la troisième démarche : elle correspond à la moyenne des écarts au carré.

Toutefois, en élevant au carré les écarts à la moyenne, on change l'unité de mesure : on ne travaille plus en euros, mais en euros au carré ce qui n'a pas de sens. C'est la raison pour laquelle on prend la racine carrée de la variance pour trouver l'écart-type. Ainsi l'unité de l'écart-type est également l'euro, comme celle des données analysées.

Fiche 5 :

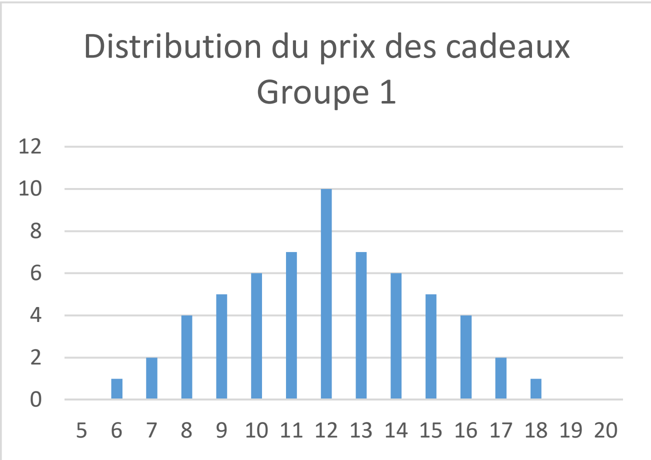
Travail en contexte de grands nombres

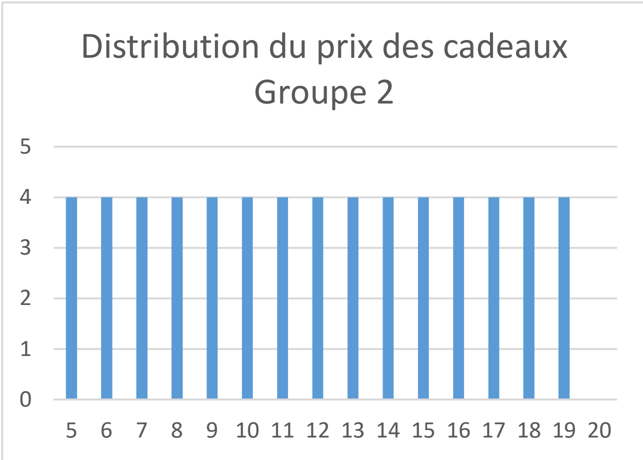
Trois groupes de 60 personnes décident de s’offrir mutuellement un cadeau : le prix de celui-ci doit être compris entre 5 et 20 €.

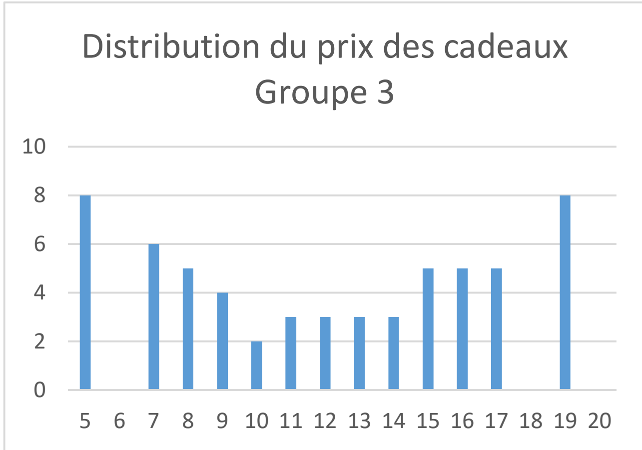
Dans les trois groupes, la moyenne est de 12€.

Tu vas à présent apprendre à calculer l’écart-type des trois distributions avec un tableur.

Avant de procéder aux calculs, observe les graphiques correspondant aux distributions.







Identifie la distribution qui a :

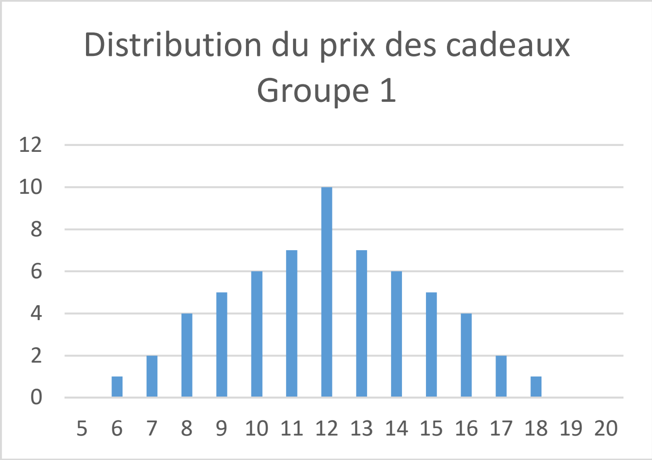
- l’écart-type le plus élevé : ……………………………………………………

- l’écart-type le moins élevé : ………………………………………………

Pourquoi élève-t-on les écarts au carré ?

Xi (sommes en euros)	ni (groupe 1)	xi.ni	xi-x (Ecart (€))	$(xi-x)^2$ (Ecart² (€²))	$ni*(xi-x)^2$
5	0				
6	1				
7	2				
8	4				
9	5				
10	6				
11	7				
12	10				
13	7				
14	6				
15	5				
16	4				
17	2				
18	1				
19	0				
20	0				
Total	**60**				

1) Complète ce tableau avec Excel.



Moyenne = ... / 60 = ...

Variance = .../ 60 = ...

Écart-type = $\sqrt{...}$ = ...

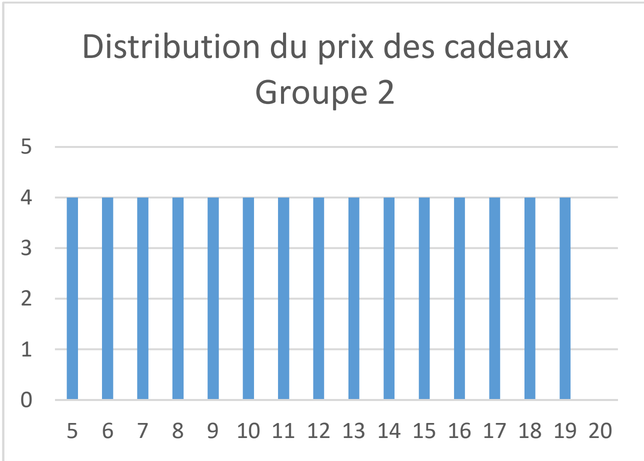
Xi (sommes en euros)	ni (groupe 2)	xi.ni	xi-x	(xi-x)²	ni*(xi-x)²
5	4				
6	4				
7	4				
8	4				
9	4				
10	4				
11	4				
12	4				
13	4				
14	4				
15	4				
16	4				
17	4				
18	4				
19	4				
20					
TOTAL	60				

Moyenne = ………/60 =

Variance = ………./60 =

Écart-type = $\sqrt{\ldots\ldots\ldots.}$ =

1) Complète ce tableau avec Excel.
2) Détermine la moyenne et l'écart-type en encodant, sur ta calculatrice, les données des deux premières colonnes du tableau.

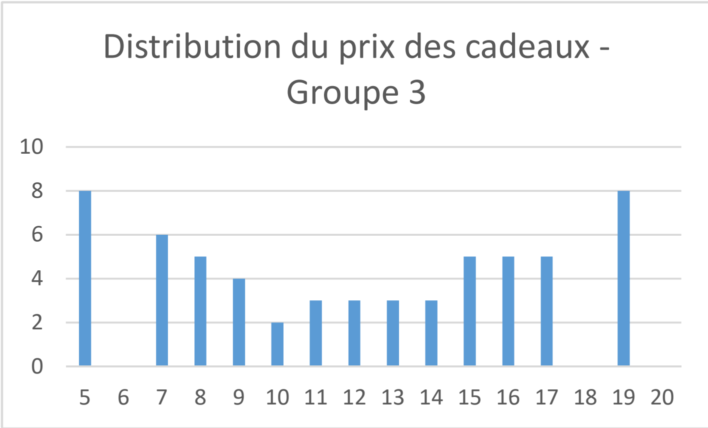


xi (sommes en euros)	ni (groupe 3)
5	8
6	0
7	6
8	5
9	4
10	2
11	3
12	3
13	3
14	3
15	5
16	5
17	5
18	0
19	8
20	
TOTAL	60

Déterminer la moyenne et l'écart-type en encodant, sur ta calculatrice, les données des deux colonnes du tableau ci-contre.

Moyenne : ……………………

Écart-type : …………………..



ACTIVITÉ 7 : DÉVELOPPER LA COMPRÉHENSION CONCEPTUELLE DE L'ÉCART-TYPE

Dans l'activité précédente, les élèves ont apris à calculer l'écart-type, en comparant différents ensembles de données dont la moyenne est chaque fois identique. Dans cette activité, les élèves analysent des distributions dont la moyenne n'est pas toujours la même. Cela leur permet de prendre conscience que l'écart-type d'une série de données doit nécessairement s'interpréter en fonction de la moyenne.

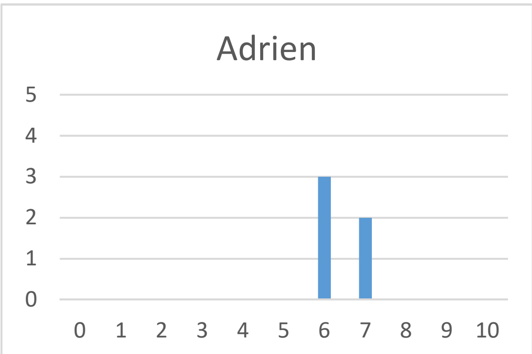
Fiche 4 :

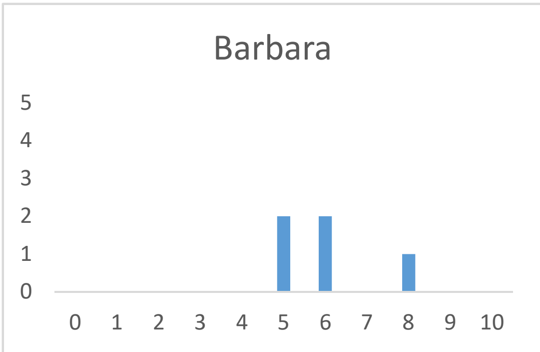
Analyser les écarts à la moyenne d'une distribution

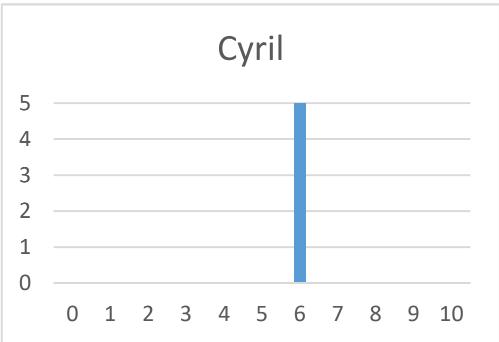
1) Cinq diagrammes montrent la répartition des notes de 5 évaluations pour les élèves suivants : Adrien, Barbara, Cyril, Denis et Emilia.

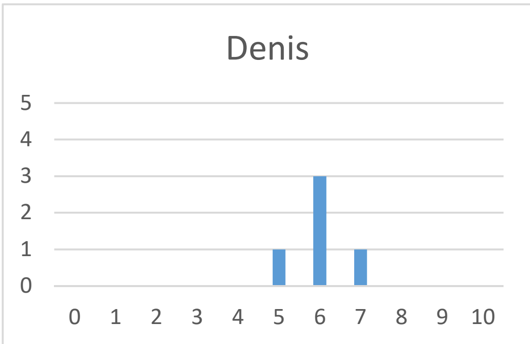
Relie chaque diagramme à une des cinq propositions ci-dessous. Justifie tes choix.

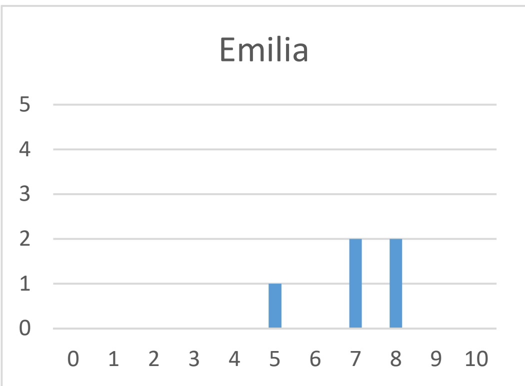
1) Moyenne = 6 Écart-type = 0	2) Moyenne = 6 Écart-type ≈ 0,63	3) Moyenne = 6 Écart-type ≈ 1,095
4) Moyenne = 7 Écart-type ≈ 1,095	5) Moyenne = 6,4 Écart-type ≈ 0,49	



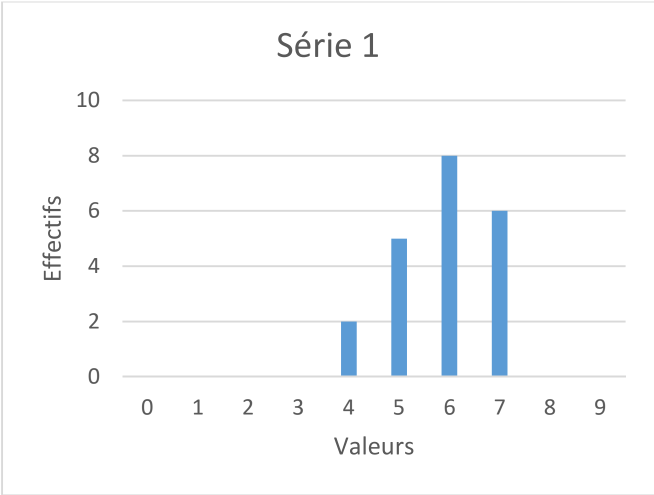


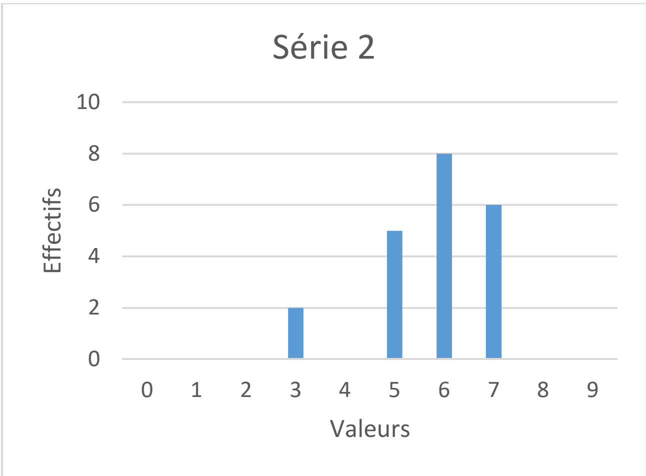






2) La position d'un seul bâtonnet diffère entre les deux diagrammes suivants.





Coche la proposition qui convient et justifie ton choix sans faire de calculs.

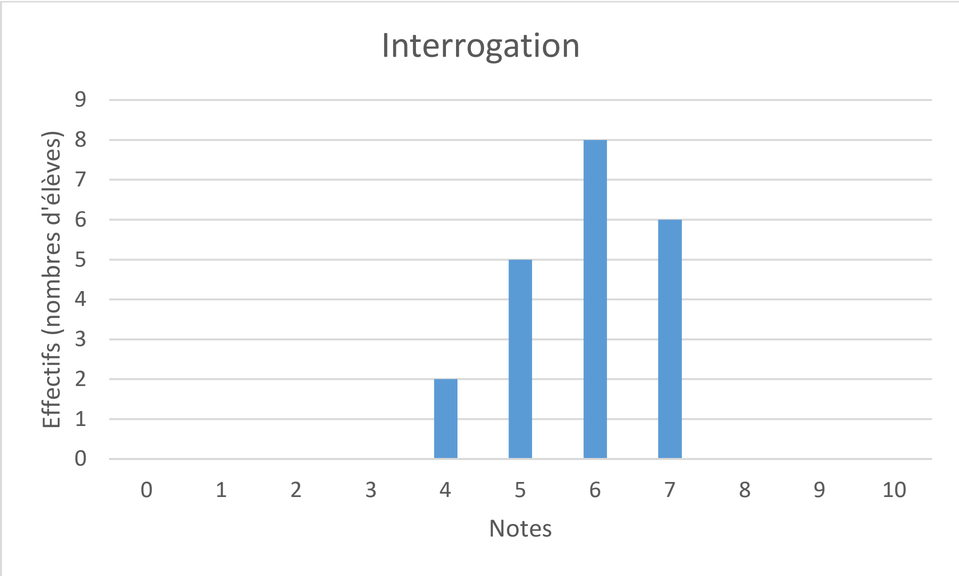
a) La moyenne de la série 2 est

- ☐ égale à la moyenne de la série 1.
- ☐ inférieure à la moyenne du diagramme 1.
- ☐ supérieure à la moyenne du diagramme 1.

b) L'écart-type de la série 2 est

- ☐ égal à l'écart-type de la série.
- ☐ inférieur à l'écart-type de la série 1.
- ☐ supérieur à l'écart-type de la série 1.

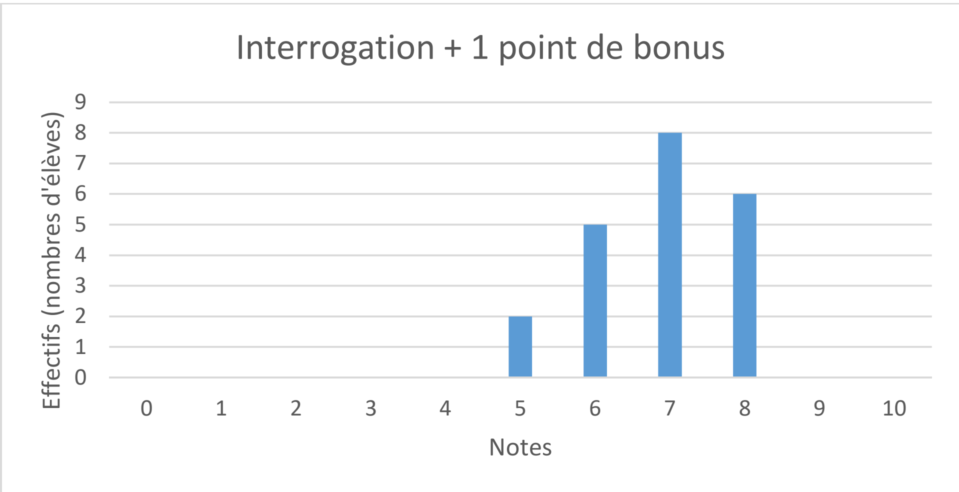
3) Le diagramme ci-dessous montre les points obtenus à un test par les élèves d'une classe.



a) Pour ce test,

- quelle est la note moyenne de la classe ?
- que vaut l'écart-type ?

b) Grâce à une question bonus réussie par tous les élèves, chaque note est augmentée de 1. Le diagramme ci-dessous montre la répartition des points des élèves en ajoutant le point bonus.

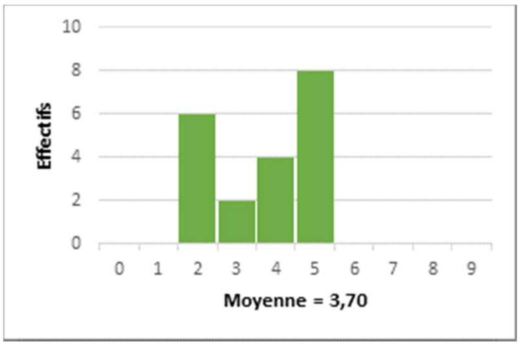


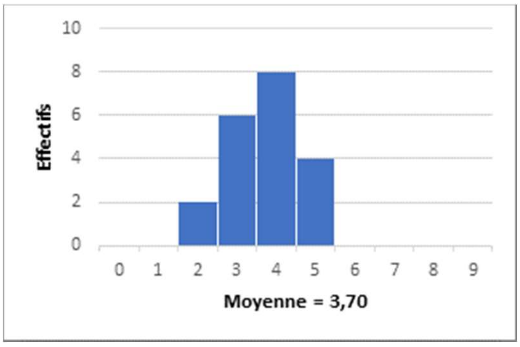
Que devient alors

- la moyenne ?
- l'écart-type pour cette nouvelle répartition ?

4) Pour lequel des deux diagrammes ci-dessous l'écart-type est-il le plus élevé ?

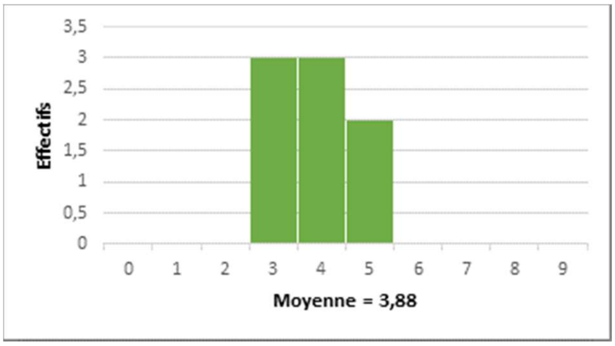
Justifie ta réponse.

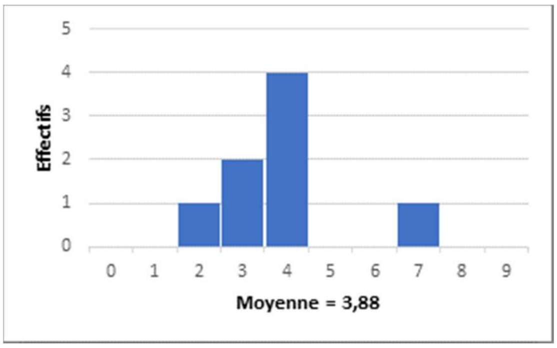




5) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

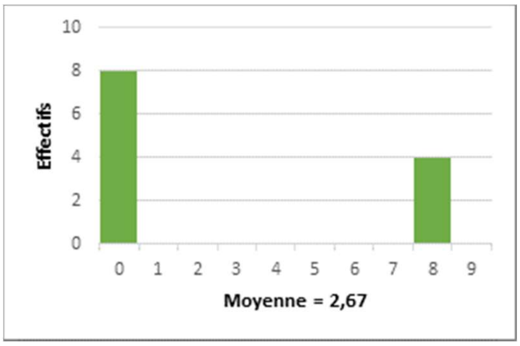
Sans calculer, quelle série possède le plus grand écart-type ? Justifie ta réponse.

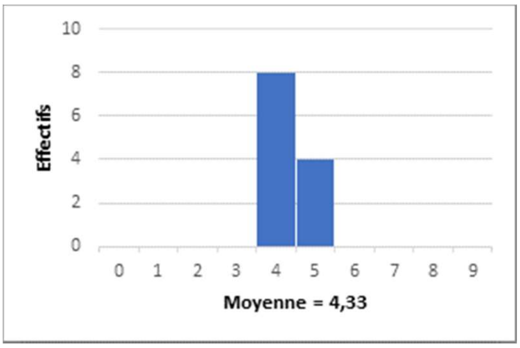




6) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

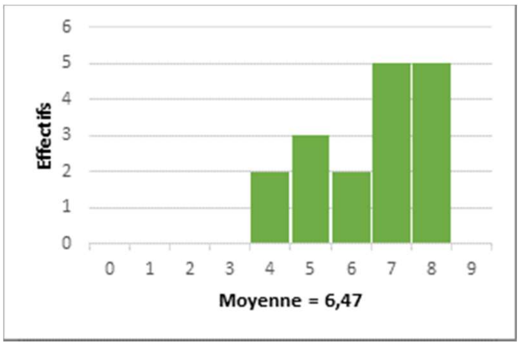
Sans calculer, quelle série possède le plus grand écart-type ? Justifie ta réponse.

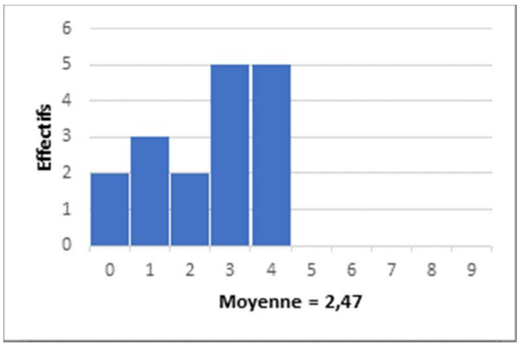




7) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

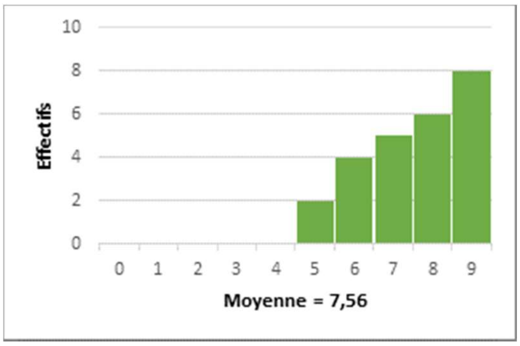
Sans calculer, quelle série possède le plus grand écart-type ? Justifie ta réponse.

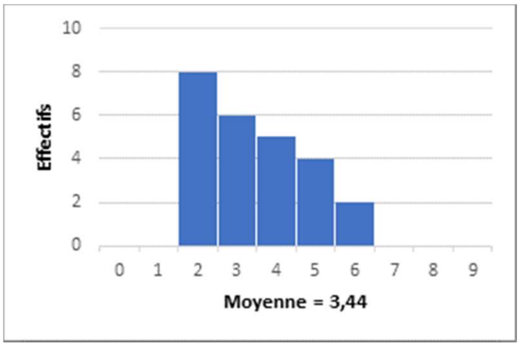




8) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

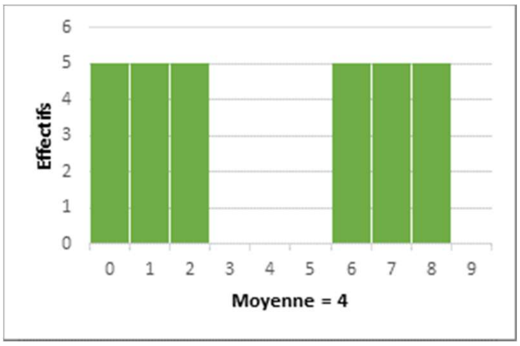
Sans calculer, quelle série possède le plus grand écart-type ? Justifie ta réponse.

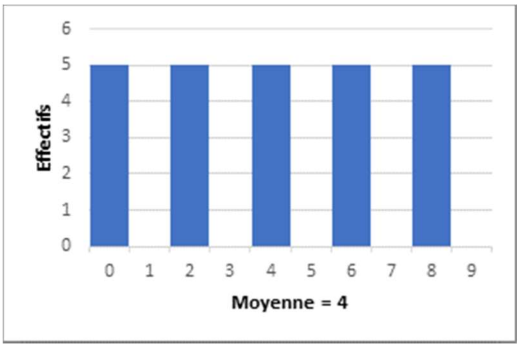




9) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

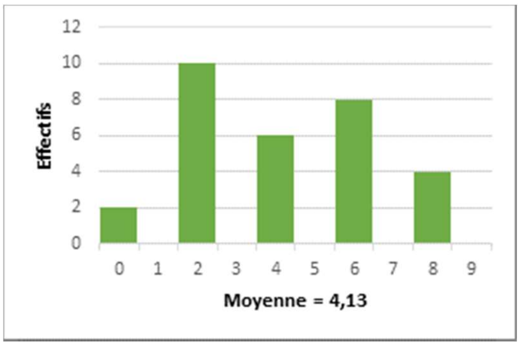
Quelle série possède l'écart-type le plus élevé ?

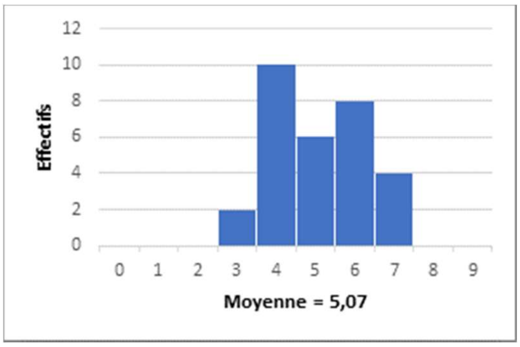




10) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

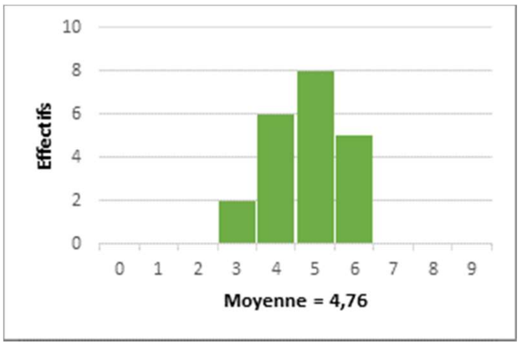
Quelle série possède l'écart-type le plus élevé ?

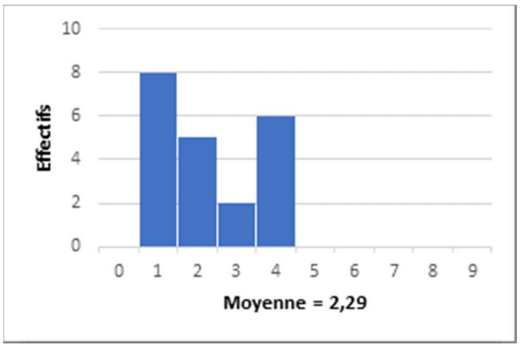




11) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

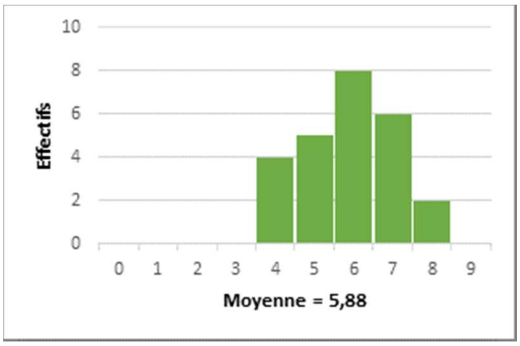
Quelle série possède l'écart-type le plus élevé ?

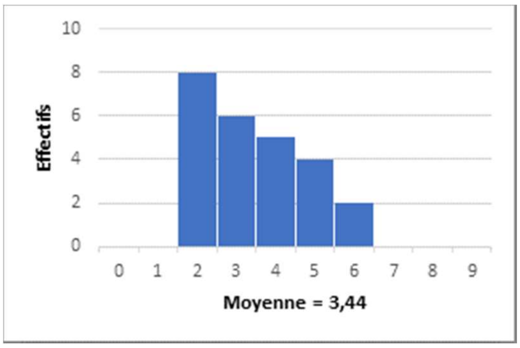




12) Voici une représentation de deux séries statistiques. La moyenne est indiquée.

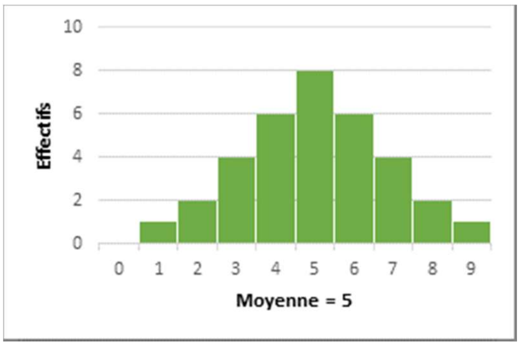
Quelle série possède l'écart-type le plus élevé ?

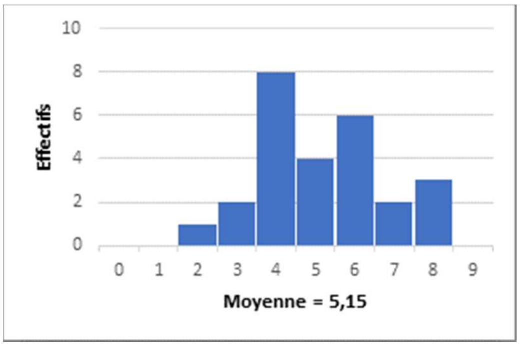




13. Voici une représentation de deux séries statistiques. La moyenne est indiquée.

    Quelle série possède l'écart-type le plus élevé ?





ACTIVITÉ 8 : QUIZZ INTERACTIF POUR REVOIR LES NOTIONS INVESTIGUÉES DANS CES PISTES DIDACTIQUES

Cette dernière activité propose un quiz interactif sur « la digitale » permettant de revoir les notions abordées dans ces pistes. Il a été créé de telle manière que l'on puisse ajouter des questions, les modifier, …

« La digitale » conçoit et développe des outils numériques pour les enseignants.

Explications de « La digitale »

Un fichier « zip » a été créé comportant l'ensemble des questions disponibles pour une activité en classe avec les élèves. Il est accessible sur www.enseignement.be/evaluationsexternes

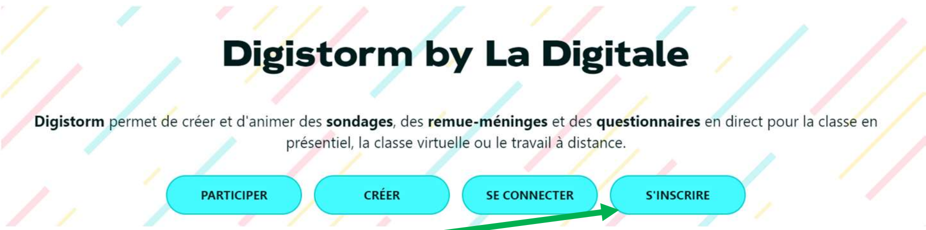
Il vous faudra, créer un compte et une fois le fichier inséré, vous pourrez le modifier, ajouter des questions ou l'utiliser tel quel en classe.

Les élèves, sans se créer de compte, pourront rejoindre le quiz en ligne et répondre aux différentes questions via un QR code, un lien ou un code. Une fois le test effectué, vous recevrez une synthèse des réponses.

Création du compte

Une fois sur le site : **https://ladigitale.dev/** , vous choisissez l'application "digistorm" et vous cliquez sur utiliser.

Création de votre compte DIGISTORM



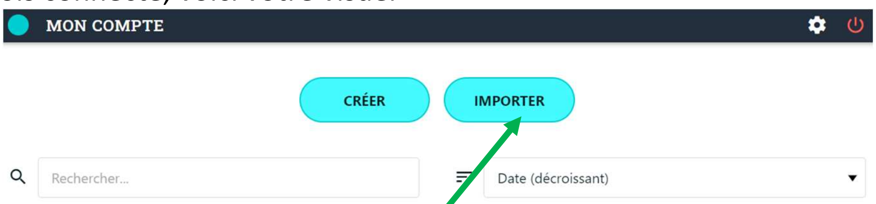
1) S'inscrire (à la première utilisation)

Après avoir cliqué sur S'INSCRIRE, il suffit de choisir votre Identifiant et votre mot de passe.

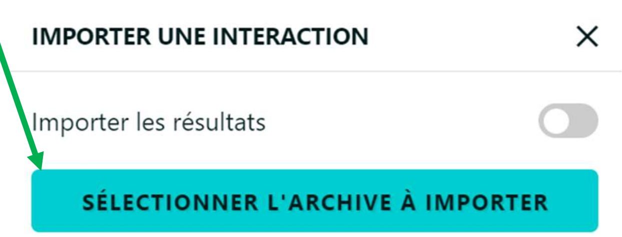
N'oubliez pas de VALIDER !

2) Se connecter : Entrez votre identifiant et votre mot de passe et sans oublier de VALIDER !

3) Une fois connecté, voici votre visuel



Il vous suffit de cliquer sur le bouton IMPORTER et puis SÉLECTIONNER L'ARCHIVE À IMPORTER



et à récupérer le fichier *NOMFICHIER*.zip à l'emplacement de votre pc où vous l'avez téléchargé.

Plus de tutos sur La Digitale sont disponibles à l'adresse https://digipad.app/p/4707/35503ca112661

RÉFÉRENCES

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, *25*(4), 527-547.

Cai, J. (2010). Exploring students' conceptual understandinf of the averaging algorithm. School Science and mathematics. Volume 98, issue 2, 93-98.

Cerclé, V. (2021). Et si on s'intéressait à la moyenne des écarts ? APMEP, 510, 398-404.

Delmas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55-82.

Garfield, J. (2002) The Challenge of Developing Statistical Reasoning, Journal of Statistics Education, 10:3.

Godino, J. D. (2002). Studying the median: a framework to analyse instructional processes in statistics education. Cape Town: International Association for Statistics Education.

Henry, V., & Miewis, J. (2016). Statistique et probabilités : les nouveautés du référentiel. Losange, n°32, 35-40.

Lappan, G. (1988). Research into Practice: Teaching Statistics: Mean, Median, and Mode. *Arithmetic Teacher*, *35*(7), 25-26.

Lawrence M. Lesser, Amy E. Wagler & Prosper Abormegah (2014) Finding a Happy Median: Another Balance Representation for Measures of Center, Journal of Statistics Education, 22:3

Mayén, S., & Díaz, C. (2010). Is median an easy concept? Semiotic analysis of an open-ended task. In *Proceedings the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Masjudin, M., Muzaki, A., Abidin, Z., & Ariyanti, I. A. P. (2020). Analysis of student's statistical thinking ability in understanding the statistical data. In *Journal of Physics: Conference Series* (Vol. 1521, No. 3, p. 032063). IOP Publishing.Sawojewski & Shaughnessy, 2000).

Mooney, E.S. (2002) A Framework for Characterizing Middle School Students' Statistical Thinking? Mathematical thinking and learning.

O'Dell, R. (2012). The mean as balance point. Mathematics teaching in the middle school. Vol 18, n°3, 148-155.

Randall E. Groth & Jennifer A. Bergner (2006) Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode, Mathematical Thinking and Learning, 8:1, 37-63, DOI: 10.1207/s15327833mtl0801_3

Schuyten, G. (1991). Statistical thinking in psychology and education. In *D. Vere-Jones (ed.), Proc. of the Third Intern. Conference on Teaching Statistics fICOTS), Dunedin, August 1990, Vol. 2, Teaching Statistics Beyond School Level, New Zealand, ISI Publications in Statistical Education* (pp. 486-490).

Uccellini, J. C. (1996). Teaching the mean meaningfully. *Mathematics Teaching in the Middle School*, *2*(2), 112-115.

Zawojewski, J., & Shaughnessy, J.M. (2000). Mathematics Teaching in the Middle School, MARCH 2000, Vol. 5, n°7. 436-440