

Introduction

Depuis quelques décennies, la statistique s'est introduite tout autant dans le monde professionnel que dans la presse quotidienne. Elle fait partie des matières de base de l'enseignement supérieur dans la plupart des disciplines et est intégrée dans les programmes de l'enseignement secondaire de transition et de qualification du D2 et/ou D3.

Qu'est-ce que la statistique ?

Une étude statistique naît d'un besoin d'information pour alimenter un processus de décision. Elle est présente dans de nombreux domaines : la recherche médicale et pharmaceutique, les tests de psychologie, les assurances... Que dire d'un monde politique privé de sondages d'opinion ? La statistique est un outil de communication universel qui permet de faire passer un message.

Pour préparer puis réaliser une étude statistique, le chargé d'étude doit disposer de suffisamment d'éléments pour établir la liste des données qu'il va essayer de recueillir au moyen d'une enquête, d'un sondage, d'un questionnaire... Il doit avoir une idée des variables statistiques à définir ainsi que du type de population à observer. A la fin de la campagne de mesures, les données vierges de tout traitement sont disponibles.

Quels sont les objectifs visés par le programme ?

La statistique descriptive, enseignée dans le secondaire, vise à construire des outils permettant d'obtenir une image « résumée » de la réalité complexe prise en compte.



Les outils statistiques synthétisent la réalité, fournissent des messages « résumés ». Un tableau, un diagramme sont des instruments de communication chiffrée que l'on voit un peu partout. Les paramètres de position et de dispersion complètent l'information. Ils résument les observations issues d'une réalité complexe selon des règles et conventions universelles.

Quels sont les outils exploités au cours de statistique ?

Outils pour ...	Présenter	Résumer	Croiser
	Tableaux	Paramètres de position	Corrélation
	Diagrammes	Paramètres de dispersion	Régression linéaire

Aujourd'hui, c'est une nécessité de maîtriser les règles de lecture et d'écriture, de comprendre et de pouvoir critiquer les chiffres qui sont abondamment servis et commentés par les médias.

Quel est le contenu de cet outil ?

Cet outil pédagogique rassemble des activités, des séquences d'apprentissage, des tutoriels d'utilisation d'un tableur qui ont servi de support à la formation : « Les statistiques ont envahi notre quotidien...Apprenons à nos élèves à être critiques ». Les activités proposées questionnent le sens de certains concepts statistiques à travers la logique de construction de la formule.

Structure

1. Interprétation de la moyenne
2. Dispersion
3. Écart-type
4. Découverte de la médiane et des quantiles
5. Calcul de la médiane
6. QCM numérique pour vérifier la compréhension des concepts
7. Application : Développer l'esprit critique
8. Utilisation d'un tableur dans le cas discret
9. Utilisation d'un tableur dans le cas continu

Chaque partie de cet outil débute par une « Fiche élèves » rassemblant quelques activités autour de la thématique ciblée qui peuvent vous aider à diversifier vos pratiques d'enseignement-apprentissage. Cette fiche est conçue pour être travaillée par l'élève en autonomie, en duo ou par groupe. Ce temps de travail doit être enrichi d'un feedback, d'un temps de structuration des acquis.

Une « Fiche prof » explicite une proposition de méthodologie ainsi que les visées pédagogiques justifiant le choix des activités.

Belle découverte,

Annick Looze, responsable de secteur.

Caractéristiques

Année	4 ^{ème} HGT
UAA	Statistique descriptive à une variable ²
Objet de l'outil	Situation d'apprentissage
Prérequis	Population – Effectifs – Fréquences – Moyenne - Etendue
Ressources mobilisées	Indicateurs de position (mode, moyenne, médiane, quartile) Indice de dispersion (étendue, variance et écart-type)

Objectifs pédagogiques

- Représenter et interpréter une information statistique
- Établir des critères visuels (boîte à moustaches) permettant de commenter des informations d'une distribution statistique, de comparer 2 distributions statistiques.
- Comprendre la logique de construction de la formule de l'écart-type.
- Développer l'esprit critique.
- Utiliser un tableur pour représenter une distribution, calculer la valeur des paramètres

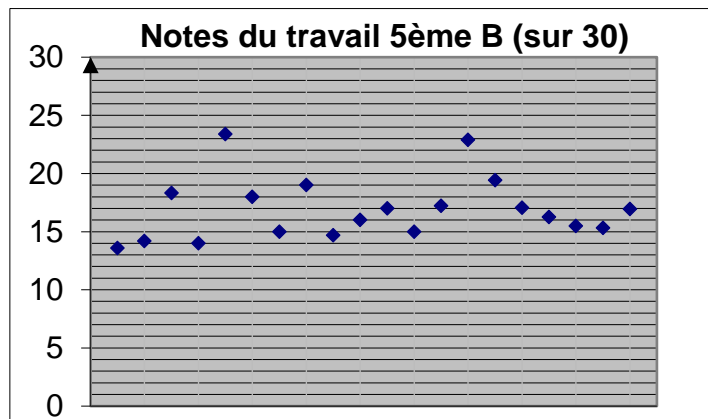
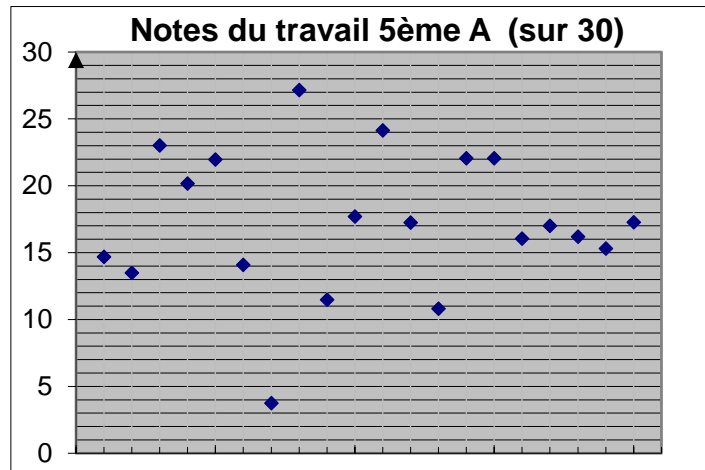
² Programme D/2014/7362/3/06

1. Interprétation de la moyenne



Énoncé 1

A l'issue de l'interrogation de géométrie en 5^{ème} A et 5^{ème} B, les 20 élèves de chacune des classes ont obtenu les résultats (sur 30) que tu peux lire sur les 2 représentations suivantes.



Sans avoir recours à un calcul, **estime** la moyenne des notes pour chacune des classes.

Énoncé 2

Imagine et représente (selon le modèle de l'énoncé 1) les notes obtenues par 6 élèves de sorte que leur moyenne sur 10, soit égale à 5.

Énoncé 3 (extrait du CE1D 2017)

QUESTION

31

/3

12 17 15 x 10

DÉTERMINE la valeur de x pour que la moyenne de ces 5 nombres soit 13.

ÉCRIS tous tes calculs.

Fiche
prof

Méthodologie

Individuellement, chaque élève répond aux questions posées dans les 3 énoncés. Une mise en commun, après chaque activité, permet d'explicitier et de confronter les différentes propositions et démarches élaborées.

Visées pédagogiques

Énoncé 1

Cette activité conduit l'élève à visualiser la moyenne comme un segment horizontal traversant un ensemble de points. Son tracé étant approximatif ($\bar{x} = 17, \dots$), plusieurs réponses sont possibles et acceptables.

Le critère selon lequel la droite a été ajustée (somme des écarts nulle) peut :

- ne pas être énoncé clairement par les élèves à cette étape (simulé seulement par un mouvement de mains) ;
- être confondu avec celui de la médiane (même nombre de points au-dessus et en dessous).

L'activité suivante permet de valider ou invalider le(s) critère(s) formulé(s) par les élèves.

Énoncé 2

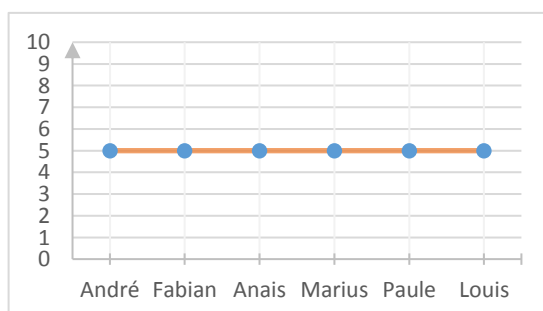
La démarche suivie pour mener à bien cette représentation sollicite l'utilisation du critère (somme des écarts nulle).

Chaque élève peut vérifier l'exactitude de sa proposition en calculant la moyenne des distributions proposées. Il valide ainsi le critère formulé ou non lors de l'activité 1.

Voici quelques représentations (non exhaustives) qui pourraient retenir l'attention :

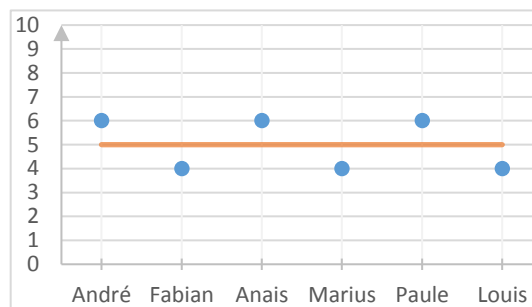
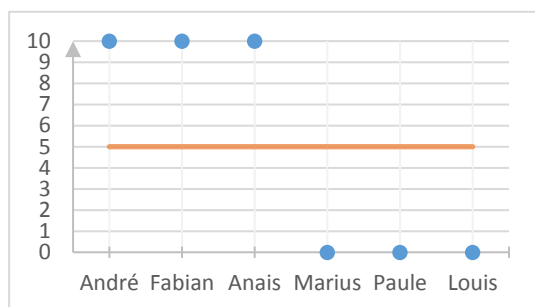
Représentation (1)

Cette représentation (1) permet de revenir sur le sens de la moyenne et de la formule associée. La note moyenne est la note qui serait obtenue par chaque élève si on suppose que chacun de ceux-ci a la même note.

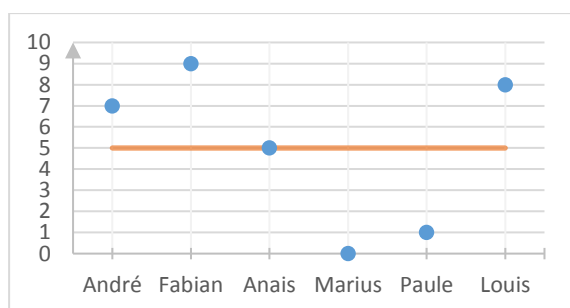


Représentations (2 et 3)

Les représentations (2) et (3) sont des graphiques possibles d'élèves qui ont formulé un critère en termes d'écart par rapport à la moyenne. Les écarts positifs sont exactement compensés par des écarts négatifs de même valeur absolue.



La représentation 4 permet de différencier la médiane et la moyenne bien que dans certaines situations (représentations 2 et 3) leurs valeurs soient identiques.

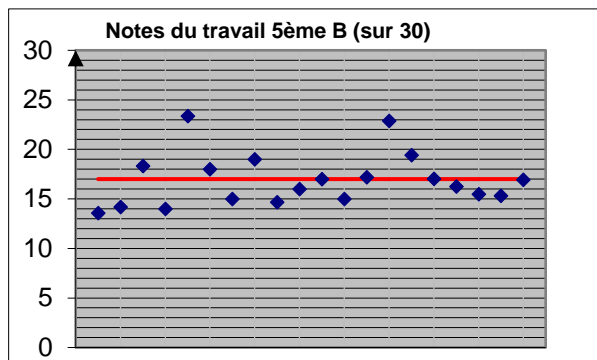
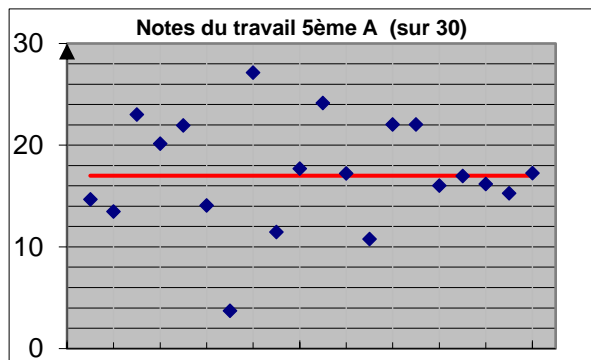
**Enoncé 3**

Cette activité permet de prendre connaissance du type de stratégies mobilisées par les élèves : représentation des données pour visualiser une « droite moyenne », compensation (somme des écarts nulle) ou résolution d'une équation.

2. Dispersion



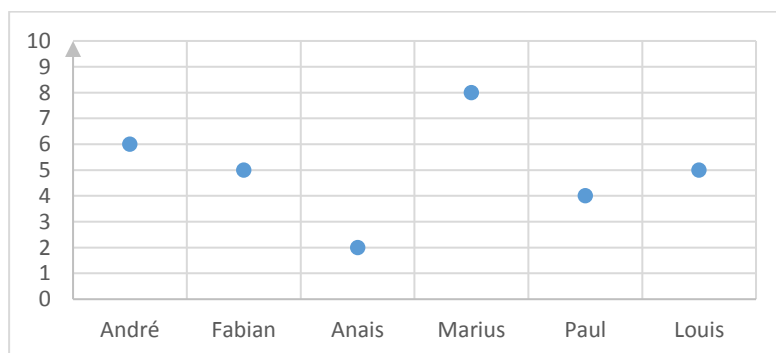
Énoncé 1



- a) Au vu des notes des classes de 5^{ème} A et B ayant obtenu la même moyenne, dans quelle classe préférerais-tu être ? Pourquoi ?
- b) Sur les 2 nuages de points de moyenne identique, les points ne sont pas distribués de la même manière autour de celle-ci. Décris avec tes mots ce qui différencie ces 2 répartitions.

Énoncé 2

Reprenons l'exemple des notes obtenues par 6 élèves dont la moyenne vaut 5/10.



Calcule une mesure de cette dispersion.

Les autres élèves ont-ils obtenu la même valeur ? Pourquoi ?

Fiche
prof

Méthodologie

Individuellement, chaque élève répond aux deux questions. Une mise en commun permet de confronter les différentes propositions et de développer des arguments pertinents justifiant le point de vue adopté.

Visées pédagogiques

Cette activité permet :

- d'introduire la notion de dispersion des points autour de la moyenne en lien avec la notion d'écart à la moyenne : « plus les écarts sont grands, plus la dispersion est forte » ;
- de motiver le besoin de définir un indicateur qui pourrait donner une mesure de cette dispersion et comparer plusieurs distributions de données.

Logique de construction de la formule

1) L'élève propose de considérer les « écarts à la moyenne ».

$$(6-5) ; (5-5) ; (2-5) ; (8-5) ; (4-5) ; (5-5)$$

$$x_i - \bar{x}$$

2) Pour mettre en relief l'idée de dispersion, il additionne tous les écarts.

$$(6-5)+(5-5)+(2-5)+(8-5)+(4-5)+(5-5) = 1+0+(-3)+3+(-1)+0 = 0$$

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

Le résultat obtenu sera nul quel que soit le cas étudié.

Les écarts qui sont au-delà de la moyenne (positifs) sont **compensés** exactement par les écarts qui sont au-dessous de la moyenne (négatifs). A cette étape, tous les élèves doivent avoir intégré cette propriété de la moyenne grâce aux activités précédentes.

3) Pour remédier au problème lié aux signes, « élever au carré tous les écarts » est un moyen, prendre la valeur absolue³ est une autre idée qui pourrait être proposée.

$$1^2+0^2+(-3)^2+3^2+(-1)^2+0^2 = 20$$

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

4) Il reste à diviser par le total des observations (n=6) pour que le résultat ainsi obtenu ne soit pas influencé par le nombre de données et donc pouvoir comparer des distributions ayant un nombre différent d'observations :

$$\frac{20}{6} = 3,33$$

³ Cfr Point attention page 9

$$\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$$

On définit la variance d'une distribution comme étant la valeur de la moyenne des écarts au carré.

5) La grandeur de départ n'est plus respectée, elle a été élevée au carré et est difficilement interprétable.

Dans notre exemple, les notes deviennent des (notes)².

Pour revenir à l'unité de départ, l'élève doit extraire la racine carrée de la variance.

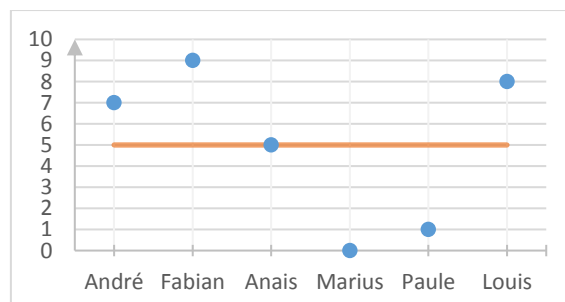
$$\sqrt{\frac{20}{6}} = \sqrt{3,333 \dots} = 1,825$$

$$\text{Ecart - type} = \sqrt{\text{Variance}} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$$

Interprétation de l'écart-type

Le professeur propose de transférer ce raisonnement à partir des notes de la représentation (4) de l'énoncé 2, partie 1.

Pour une même moyenne (5/10), la valeur de l'écart-type est de 3,4 points.



Comme nous disposons des données, nous pouvons dire de manière précise que 3 élèves ont une note comprise dans l'intervalle $[1,6; 8,4]$ c'est-à-dire $[\bar{x} - \sigma; \bar{x} + \sigma]$.

Dans le cas où les données ne sont pas connues ou leur nombre trop élevé, l'inégalité de Tchebychev⁴ est un des outils d'interprétation de l'écart-type.

Le calcul de l'écart-type prend tout son sens lorsqu'il s'agit de **comparer la dispersion** de 2 populations pour une même variable.

Point d'attention

Prendre la valeur absolue des écarts à la moyenne est une autre stratégie pour écarter le signe des écarts. Les écarts absolus moyens par rapport à la moyenne arithmétique se prêtent mal aux calculs algébriques. On leur préfère donc l'écart-type. Néanmoins, l'étude de la moyenne des écarts absolus conduit à un autre paramètre de position : la médiane qui minimise cette somme. Ce point de théorie est développé dans un article de l'APMEP téléchargeable en activant le lien suivant : [cliquer](#)

⁴ Voir article « Statistique et probabilités : les nouveautés du référentiel » extrait de la revue Losanges N°32 2016 SBPM [site web]. Consulté le 14 mai 2019.
<https://www.sbpn.be/losanges/>

3. Découverte de la médiane et des quantiles



Enoncés

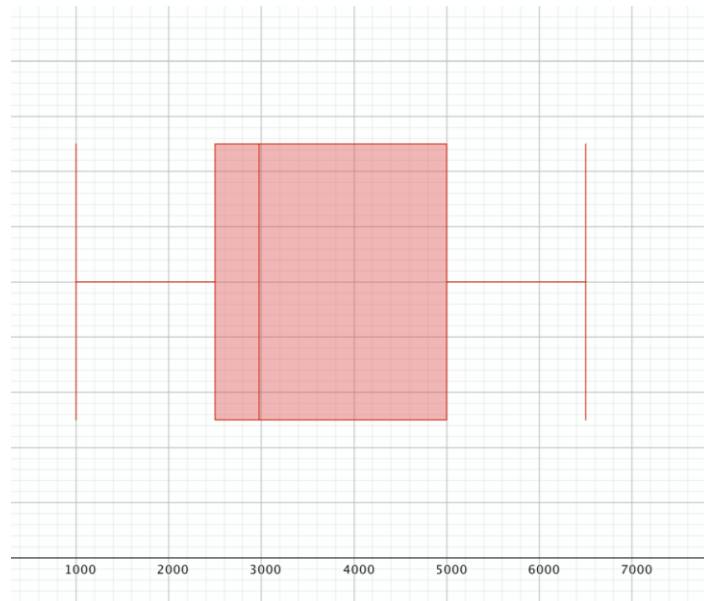


Pourquoi gagnez-vous moins que le salaire moyen ?

<https://www.youtube.com/watch?v=ulx2xvdwll0>

1. A partir de cette vidéo, relève le vocabulaire qui te semble utile pour la compréhension de son contenu.
2. En Belgique, connais-tu l'organisme qui récolte les données relatives aux salaires ? Sur quel site internet peux-tu les trouver ?
3. Prends connaissance des questions suivantes et réponds-y à partir des informations données à travers la vidéo.
 - a) Vérifie par calcul la valeur des 2 moyennes annoncées dans la vidéo.
 - b) « *La moyenne de Albert à Max est de 2800 euros. Bien qu'aucun des 4 chats n'ait gagné 2800 euros, quelqu'un qui tenterait de résumer la situation des 4 chats par cette moyenne ne se tromperait pas.* »
Explique pourquoi et ce que cette valeur signifie.
 - c) « *Si on utilise la valeur de 3540 euros comme moyenne, l'image ne colle pas à la réalité.* »
Quels outils utilisent alors les statisticiens pour se rapprocher de la réalité ?
 - d) Dans la vidéo, on parle de 4 quintiles et de 5 blocs ; de 9 déciles et de 10 blocs.
Qu'en serait-il s'il s'agissait de quartiles ?
 - e) La valeur du 5^{ème} décile est de 2976 euros. Quel autre nom donne-t-on à ce paramètre ? Interprète le résultat.

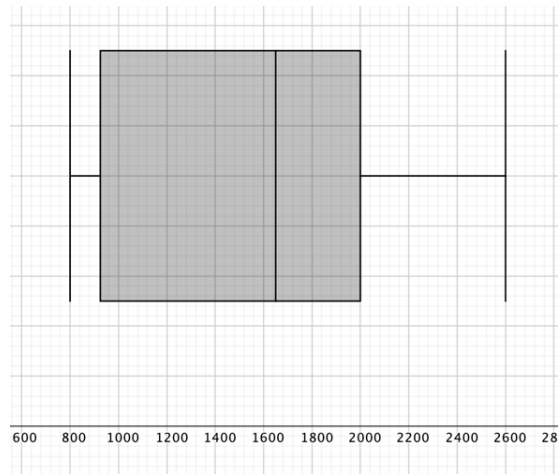
4. Une représentation couramment utilisée en statistiques s'appelle « Boîte à moustaches » - BAM . Voici celle des salaires des chats de la vidéo :



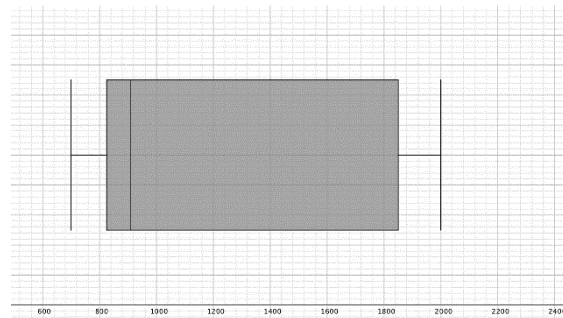
- Place sur ce diagramme la valeur de la médiane (2976 €), du premier quartile (2500 €) et du troisième quartile (5000 €).
- Que désigne le premier trait vertical de la BAM ? Et le dernier ?
Que représente la distance entre ces 2 traits ?
- Place sur ce diagramme le salaire d'Albert. Peut-on dire qu'au moins 25% des personnes gagnent moins que lui ?
- L'intervalle interquartile** d'une distribution statistique est un intervalle dont les bornes sont le 1^{er} et le 3^{ème} quartile.
 Quel pourcentage de la population comprend cet intervalle ?
 Quelles sont les bornes de cet intervalle dans ce cas ?
 Complète la phrase suivante qui reprend ces informations :
 « % de personnes perçoivent des salaires compris entre euros et euros ».

5. Compare les deux boîtes à moustaches ci-dessous, illustrant la répartition des salaires mensuels (euros) des ouvriers de deux entreprises.

Entreprise 1

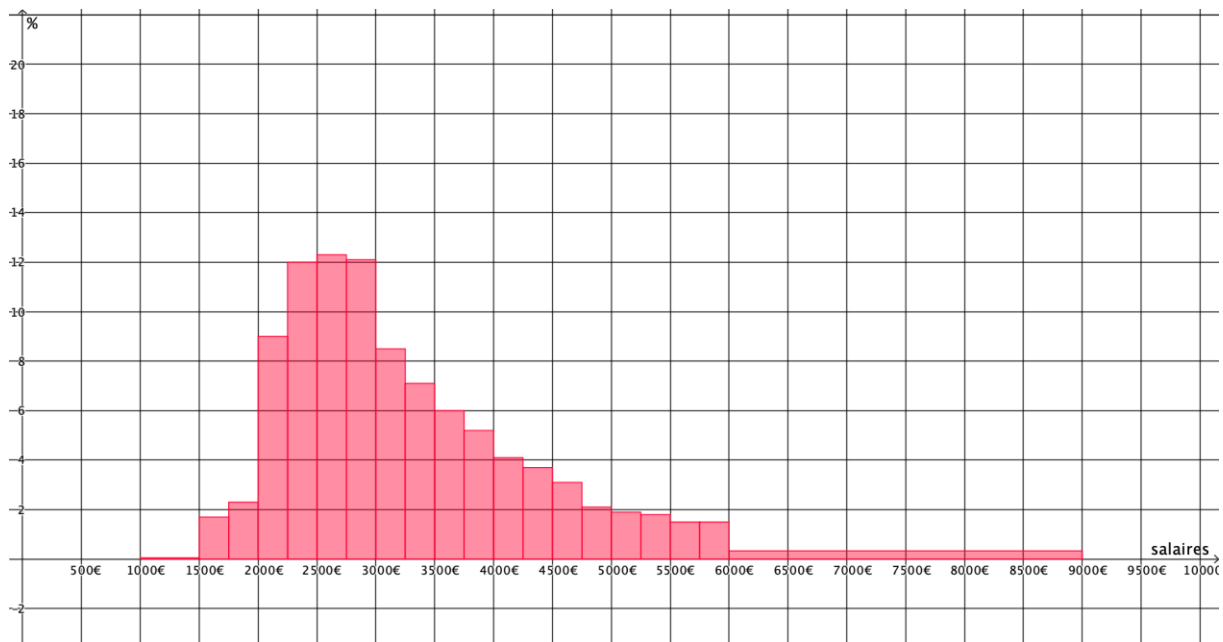


Entreprise 2



6. Voici les histogrammes représentant la distribution des salaires belges bruts mensuels (euros) de 2015, réalisés à partir des données disponibles sur STATBEL5 (site officiel de statistiques belge).

Distribution des salaires bruts des femmes

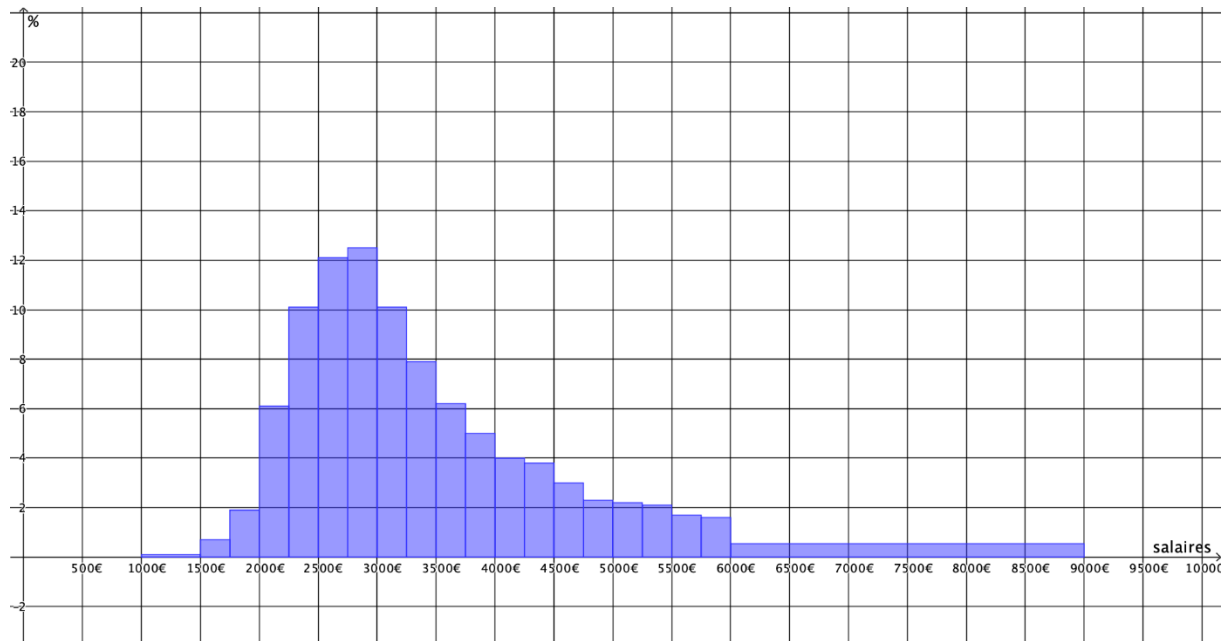


⁵ STATBEL [site Web].

Consulté le 30 mars 2019.

<https://statbel.fgov.be/fr>

Distribution des salaires bruts des hommes



- Quel pourcentage approximatif de femmes gagne au plus 1999,99 euros ?
- Quel pourcentage approximatif d'hommes gagne au moins 5500 euros ?
- Construis la BAM des hommes en dessous du diagramme statistique
- Commente à partir de cette distribution la phrase extraite de la vidéo :
« Les femmes continuent à gagner moins que les hommes ».



Méthodologie

Le professeur organise l'activité en classe informatique ou en classe de math avec le matériel dont il dispose (smartphone, projecteur, tablette, ...).

- A la suite du **premier point**, il est important que l'enseignant revienne sur le vocabulaire épinglé par les élèves et en garde une trace écrite (sous forme de carte mentale, tableau structuré...).

Salaire net – Salaire brut – Distribution – Moyenne – Quintile – Décile – Médiane

- **Avant d'aborder le point 4**, en s'appuyant sur la trace écrite réalisée au point 1, une restructuration des indices de position (médiane, quartile, intervalle interquartile, ...) semble nécessaire. Les réponses aux différentes questions permettent d'épingler certaines caractéristiques de la moyenne et de la médiane :
 - la moyenne est un paramètre sensible et instable par rapport à la présence de valeurs extrêmes.
 - le calcul de la médiane semble judicieux dans les investigations où l'idée de classement permet une meilleure compréhension de la situation.
- **A la suite de la question 4**, le professeur dispose de tous les éléments pour construire une boîte à moustaches avec ses élèves.
- Le **point 5** permet l'interprétation de la position et de la dispersion des 2 distributions statistiques au travers de la comparaison de leurs boîtes à moustaches. Ce sera aussi l'occasion de relever la dissymétrie des boîtes à moustache. Lorsqu'on souhaite comparer 2 distributions à l'aide des BAM, il est évidemment nécessaire d'utiliser la même graduation.
- Le **point 6** présente une répartition des salaires inspirée des données collectées par un organisme officiel [STATBEL](https://statbel.fgov.be/fr)⁶. Ce type de diagramme fait référence à des données réparties en classes. Il nous semble intéressant de le présenter pour donner un exemple de variable statistique quantitative continue⁷ et pour clôturer l'activité avec un diagramme issu de la vie réelle. Le commentaire du point 6C peut s'appuyer sur la comparaison des BAM des hommes et des femmes.

Il peut arriver que la distribution contienne des données éloignées, voire aberrantes. Un critère empirique de repérage de ce type de données est explicité dans l'article⁸ extrait de la revue Losange N°32 éditée par la SBPM. Si c'est le cas, la boîte à moustaches subira quelques modifications⁹. Une valeur trop extrême sera identifiée avec une croix au-delà des moustaches.

⁶ STATBEL [site Web]. Consulté le 14 mai 2019.

<https://statbel.fgov.be/fr>

⁷ Sur le site de STATBEL, on trouve un diagramme en batônnetts. Les classes salariales y sont envisagées comme un caractère qualitatif.

⁸ SBPM [site web]. Consulté le 14 mai 2019.

<https://www.sbp.be/losanges/>

⁹ Exemple sur Alloprof [site Web]. Consulté le 14 mai 2019.

<http://www.alloprof.qc.ca/BV/pages/m1366.aspx>

Visées Pédagogiques :

- Donner du sens aux apprentissages en travaillant dans une situation contextualisée
- Utiliser les trois axes essentiels en didactique : explorer, extraire, expliquer
- Modéliser par une boîte à moustaches une situation réelle

La vidéo permet :

- a) à chaque élève de travailler à son propre rythme ;
- b) des arrêts sur image pour travailler les concepts importants ;
- c) une approche ludique et dynamique ;
- d) l'utilisation de nouvelles technologies de l'information et de la communication.

Le diagramme proposé au point 6 permet le transfert du travail réalisé à partir d'une vidéo reprenant un vocabulaire vulgarisé et des données restreintes à une situation à l'échelon national avec des valeurs réelles.

Si l'enseignant souhaite développer la compétence « lire un document scientifique » avec un vocabulaire adapté, il peut consulter la page du site [STATBEL](#) sur laquelle on peut lire certaines informations reprises dans la vidéo :

4. Calcul de la médiane et des quantiles

Cette partie reprend différentes procédures de calcul de la médiane en fonction de la manière dont les informations sont fournies : une liste de données, des données groupées (tableau de distribution) ou le graphique des effectifs (fréquences) cumulés(es). Dans le cas particulier de la variable continue, nous avons privilégié l'approche par interpolation linéaire. Elle permet d'explicitier la signification de cette expression ainsi que l'hypothèse de départ qui sous-tend cette technique, de mobiliser des notions déjà acquises, de s'appuyer sur un support visuel qui évite la mémorisation d'une formule « vide de sens ».

Dans certaines investigations, ce sont les idées de « plus de, moins de », « d'ordre », de « classement » qui sont adaptées à la compréhension de ces situations. La médiane et les quantiles s'avèrent dans ces cas, les paramètres statistiques les plus adéquats.

Pour calculer la médiane, on veillera à varier le type d'informations : fréquences cumulées ou effectifs cumulés.

La **médiane** d'une série statistique est la valeur qui sépare cette série ordonnée en 2 sous-séries d'effectifs égaux.

Le calcul de la médiane est réalisé de façon différenciée en fonction de la présentation des données :

4.1. Liste de données

Si le nombre d'observations est **impair**, la médiane est toujours parfaitement déterminée :

{3,6,12, **15**, 21,28,32} → la médiane est 15

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

{3,6,12, **15, 21**, 28,32,38} → L'intervalle médian est délimité par les modalités 15 et 21.

Dans cette situation, la valeur de la médiane est, par convention, le milieu de l'intervalle : $\frac{15+21}{2} = 18$

4.2 Données groupées (tableau de distribution)

Exemple : Distribution groupée du nombre d'enfants par famille

Nbre d'enfants	Nbre de ménages	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. Tant qu'on ne considère que les modalités 0 et 1, seuls 42,5% de la population sont pris en compte. Il faut y ajouter une certaine partie des observations ayant pour valeur « 2 » pour atteindre 50 % des effectifs. La médiane, dans ce cas, vaut donc 2.

4.3 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
	[1100,1200[1150	59
classe médiane →	[1200,1300[1250	123
	[1300,1400[1350	130
	[1400,1500[1450	140
		N=140 N/2=70	

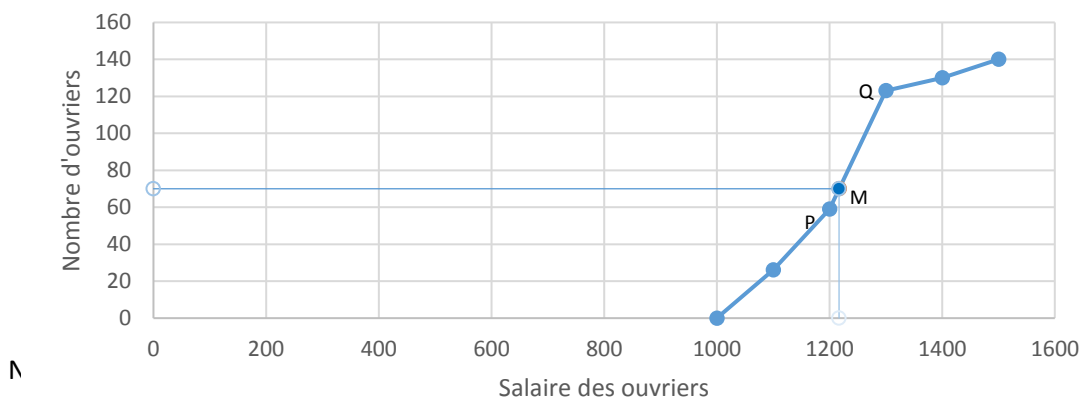
La valeur de la médiane se trouve à l'intérieur de la classe [1200,1300[

Comment calculer la valeur exacte de la médiane ?

- 2) Calculer les effectifs cumulés
- 3) Repérer la classe médiane
- 4) 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).

Polygone des effectifs cumulés



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...

6. Evaluer la compréhension des concepts



Vrai ou Faux

Pour chaque affirmation décrite ci-dessous, cochez la case correspondant à votre choix et corrigez la proposition si elle est erronée.

Faux

Vrai

		Faux	Vrai
1.	La moyenne arithmétique d'une distribution n'est pas nécessairement un élément (une modalité) de cette distribution.	<input type="checkbox"/>	<input type="checkbox"/>
2.	Une même distribution statistique peut avoir plusieurs moyennes arithmétiques différentes.	<input type="checkbox"/>	<input type="checkbox"/>
3.	A chaque individu d'une population, on peut associer une ou plusieurs modalités du caractère.	<input type="checkbox"/>	<input type="checkbox"/>
4.	Environ 75% des valeurs d'une distribution ordonnée se situent dans l'intervalle interquartile.	<input type="checkbox"/>	<input type="checkbox"/>
5.	L'écart interquartile est plus ou moins influencé par les valeurs extrêmes.	<input type="checkbox"/>	<input type="checkbox"/>
6.	On peut rencontrer des distributions statistiques possédant plusieurs modes.	<input type="checkbox"/>	<input type="checkbox"/>
7.	L'écart-type a une valeur toujours positive.	<input type="checkbox"/>	<input type="checkbox"/>
8.	$\sum_{j=1}^n (x_j - \bar{x}) = 1$	<input type="checkbox"/>	<input type="checkbox"/>
9.	Déterminer la nature des caractères suivants : a) le sexe qualitatif <input type="checkbox"/> quantitatif <input type="checkbox"/> b) la tension artérielle qualitatif <input type="checkbox"/> quantitatif <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	Le signe de la variance dépend des valeurs prises par la variable statistique.	<input type="checkbox"/>	<input type="checkbox"/>
11.	La moyenne est un nombre toujours positif.	<input type="checkbox"/>	<input type="checkbox"/>
12.	Le mode est la valeur de l'effectif maximal.	<input type="checkbox"/>	<input type="checkbox"/>
13.	La moyenne est influencée par des valeurs extrêmes d'une distribution de données	<input type="checkbox"/>	<input type="checkbox"/>



Une version numérique et interactive de ce QCM est téléchargeable en activant le lien suivant : **Kahoot!**

7. Applications – Développer l'esprit critique



Énoncés

1. Que penses-tu des affirmations suivantes :
 - a) Un responsable d'un mouvement de jeunesse prépare ses vacances d'été à la montagne ; il projette d'effectuer avec ses jeunes une balade pédestre qui nécessitera des chaussures adéquates qui seront louées sur place. Pour réserver les chaussures, il relève la pointure de ceux qui vont partir avec lui et transmet à l'organisateur local la moyenne arithmétique de ces pointures
 - b) Pour avoir une bonne idée des salaires perçus par les travailleurs d'une entreprise, il suffit de calculer le salaire moyen.
 - c) La température moyenne de la journée est de 15° . Je peux donc partir le matin de chez moi sans manteau.
 - d) Je vais chez le médecin et celui-ci me demande si j'ai fait de la température. Je lui réponds qu'après plusieurs relevés sur deux jours, j'ai une température moyenne de 38° . Cette valeur est-elle représentative pour le médecin ?

2. Petite histoire de moyenne ...

(Inspiré du livre « Le secret des paradoxes » de M. Gardner)

Après avoir lu attentivement le texte ci-dessous, relève les arguments qui explicitent, dans ce contexte, que la moyenne n'est pas un paramètre pertinent.

La situation se passe dans la société Perculum, fabricant de percolateurs.

La direction de l'usine est composée de Mr Perculum lui-même, son frère et 6 parents proches. Le personnel comporte 5 contremaîtres et 10 ouvriers. Les affaires sont florissantes et il faut engager un nouvel ouvrier. Maxime se présente à l'entretien d'embauche... Les conversations suivantes ont été entendues...

J'aimerais beaucoup travailler au sein de votre entreprise, Mr Perculum, dit Maxime. On m'a dit que les salaires y étaient attrayants !

En effet, mon garçon... Ici les salaires sont élevés. Le salaire moyen est de 1500 € par mois... Pendant votre période d'essai, vous serez payé 375 €, mais votre salaire évoluera vite.

Après quelques jours, Maxime, intrigué par les dires de son patron, désire le rencontrer.

J'aimerais vous parler Mr Perculum... Vous m'avez berné ! J'ai interrogé d'autres ouvriers de l'entreprise, et aucun d'eux ne gagne plus de 500 € par mois. Comment le salaire moyen pourrait-il être de 1500 € par mois ?

Ne vous énervez pas Maxime, le salaire moyen est effectivement de 1500 € par mois et je vais vous le prouver. Je gagne 12000 € par mois, mon frère en gagne 5000, mes 6 parents 1250 chacun, les 5 contremaîtres 1000 € et les 10 ouvriers gagnent 500 € par mois, ce qui fait un total de 34500 € par mois pour 23 personnes ; d'accord ?

Tout à fait d'accord pour ces calculs. Cela fait une moyenne de 1500 € par mois mais vous m'avez quand même berné !

Point du tout ! Vous n'avez pas compris... J'aurais pu établir la liste des salaires par valeurs croissantes et vous dire que le salaire central est de 1000 € par mois mais ce n'est pas la moyenne, c'est la médiane!

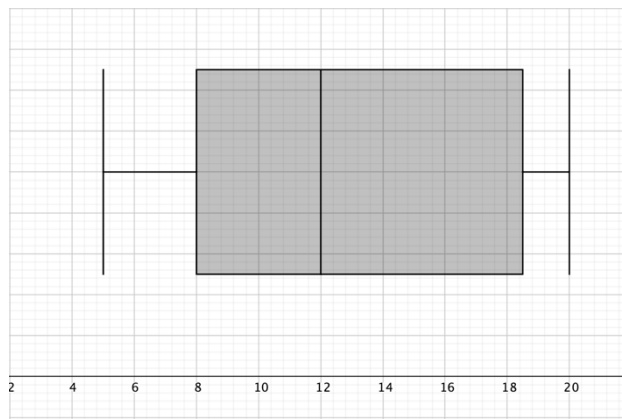
Et mes maigres 500 par mois ?

C'est le mode ! C'est-à-dire le salaire que la plupart des gens touchent effectivement. Votre problème, mon cher garçon, est que vous ignorez totalement la différence qu'il y a entre le mode, la médiane et la moyenne...

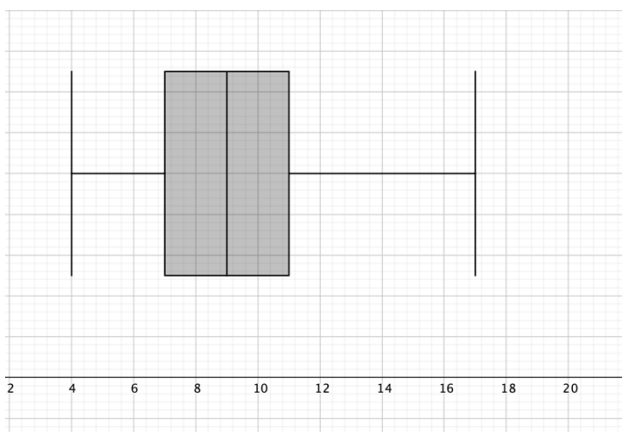
Eh bien, je la connais maintenant... mais je démissionne !

3. D'après les informations contenues dans les boîtes à moustaches, commente les propositions suivantes.
- En 4C, le quart des élèves ont moins de 7/20.
 - En 4F, la moitié des élèves ont entre 8 et 12.
 - En 4F, il y a à peu près autant d'élèves qui ont entre 6 et 10 que d'élèves qui ont entre 12 et 18.
 - En 4C, il y a à peu près autant d'élèves qui ont entre 7 et 9 que d'élèves qui ont entre 9 et 11.
 - S'il y a 28 élèves en classe de 4C, combien d'élèves ont plus que la moyenne (9/20) ?
Combien d'élèves ont moins que 11 ?

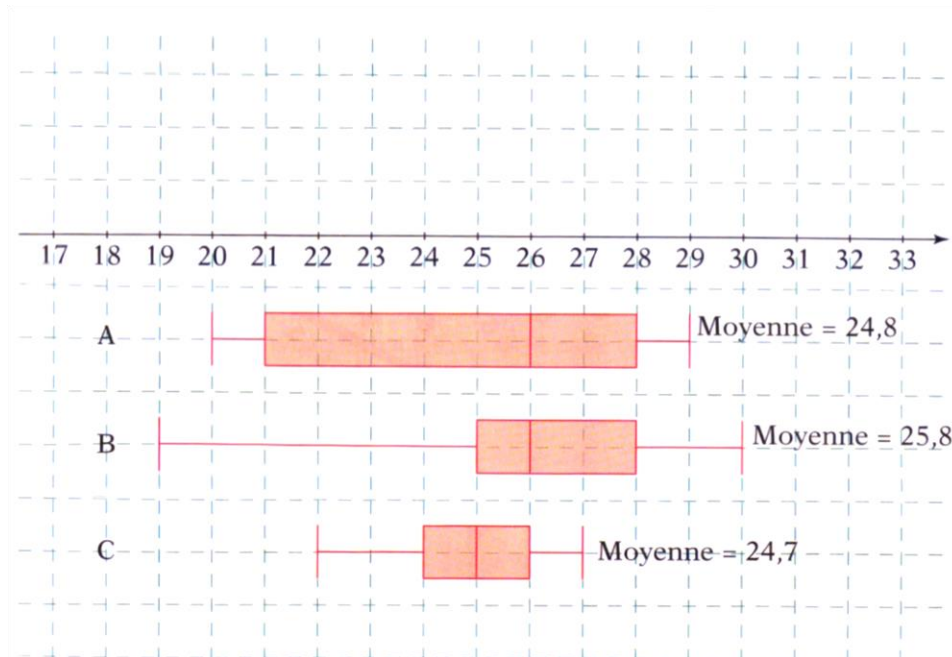
4^{ème} F



4^{ème} C



4. Ci-dessous, les diagrammes en boîtes représentant la température en mai dans les trois villes A, B et C sont donnés.
Quelle ville choisir pour y passer ses vacances ?
Plusieurs réponses sont possibles mais il faut justifier le choix en explicitant les critères et leur lien avec les valeurs statistiques fournies.



Extrait du manuel CQFD 4^{ème}



Méthodologie

Chaque élève répond aux questions. Une mise en commun permet de confronter les différentes propositions et de développer des arguments pertinents justifiant le point de vue adopté.

Visées pédagogiques

- Les énoncés 1 et 2 permettent de prouver que la moyenne arithmétique n'est pas toujours intéressante, qu'elle peut être influencée par des valeurs extrêmes et donc ne plus être représentative. Il paraît donc judicieux de mettre en relation les valeurs de la moyenne et de la médiane afin d'affiner la description de la distribution.
- L'exercice 3 permet une lecture des informations transmises par une boîte à moustaches. Ici, l'enseignant montrera que l'intérêt de cet outil est de présenter visuellement une synthèse des informations utiles à l'interprétation d'une situation.
- L'exercice 4 permet de donner du sens aux BAM au travers de leur comparaison. Il est important de tracer les boîtes à moustaches avec la même échelle et de les placer les unes en dessous des autres (ou les unes à côté des autres). On peut aussi insister sur la différence d'informations que fournit la moyenne ou la médiane.
- A partir du moment où ce chapitre a été abordé avec les élèves, il peut être intéressant que le professeur s'appuie sur ce modèle de boîte à moustaches pour communiquer les résultats des interrogations des élèves. Ainsi, chacun peut situer ses notes par rapport à celles de l'ensemble de la classe.

8. Utilisation d'un tableur dans le cas discret (GeoGebra)

Comme le programme préconise le recours à la calculatrice ou à des logiciels appropriés (les tableurs) aussi systématiquement que possible, nous vous proposons ci-dessous une activité possible.



Énoncé

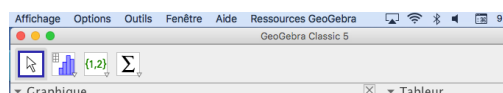
Pour un ensemble donné de 100 joueurs de basket de la NBA (stats.nba.com), le nombre de paniers à 3 points varie de 32 à 38. On obtient le tableau suivant :

Paniers à 3 points	32	33	34	35	36	37	38
Nombre de joueurs	2	3	20	25	22	20	8

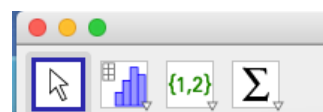
- Déterminer le nombre moyen de paniers à 3 points marqués sur une saison de la NBA
- Déterminer l'écart type
- Déterminer la médiane, les quartiles
- Tracer la boîte à moustaches

Utilisation du logiciel GeoGebra

1. Fermer la fenêtre algèbre et ouvrir la fenêtre Tableur (Menu Affichage : Cocher Tableur et décocher Algèbre)
2. Entrer dans les colonnes du tableur les valeurs des modalités et des effectifs (Colonnes A et B). Comme dans Excel, pour la colonne des modalités, on peut entrer manuellement les deux premières valeurs et tirer sur le petit carré bleu en bas à droite pour continuer la suite arithmétique.
3. Sélectionner la colonne des modalités et cliquer sur la deuxième icône (l'icône bleue avec les graphiques) et sélectionner Statistiques à une variable.



	A	B	C
1	32	2	
2	33	3	
3	34	20	
4	35	25	
5	36	22	
6	37	20	
7	38	8	
8			
9			
10			



4. Une nouvelle fenêtre apparaît. Cette petite fenêtre comporte une roue dentée à droite pour choisir le type de données. Choisir ici « Données avec effectifs ».

5. La liste des modalités s'insère alors dans la première colonne (A1 :A7).

6. Sélectionner la colonne B - Cliquer sur la petite main (à côté du mot Effectifs) . Les valeurs de la colonne B apparaissent sous la colonne des effectifs.

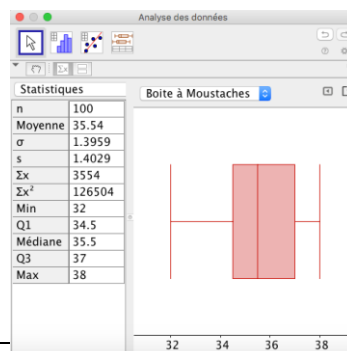
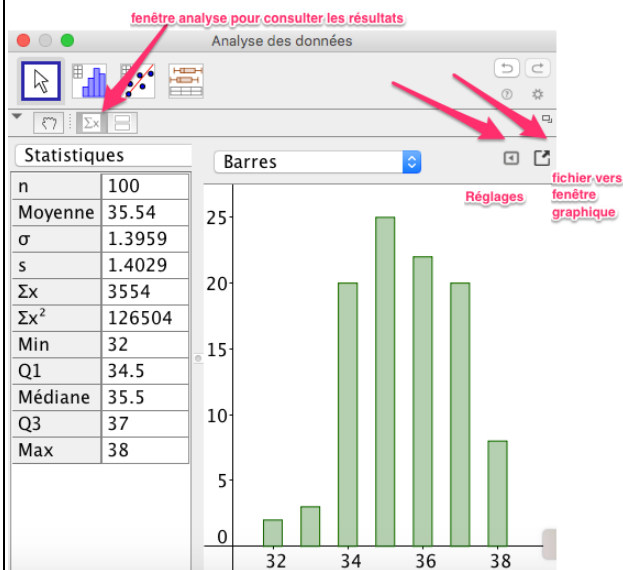
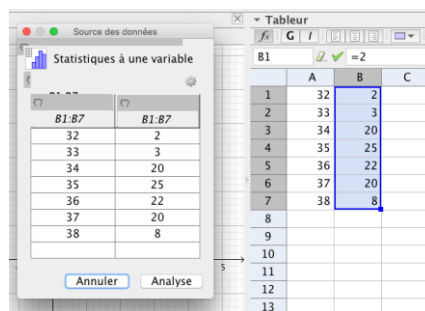
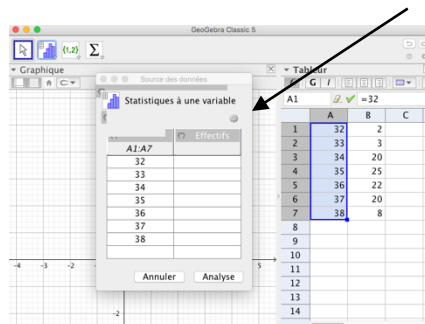
7. Cliquer sur « Analyser » (sigle Σx), pour obtenir les valeurs caractéristiques de la distribution.

(moyenne, médiane, quartiles, écart-type).

8. Un diagramme en barres est tracé à partir des valeurs des deux colonnes données. On peut changer les largeurs des bâtonnets et obtenir le diagramme dans la fenêtre graphique de GGB si on le souhaite en cliquant sur la petite icône à droite du graphique.

N.B. GGB utilise la définition (* page 16) de la médiane dans le cas discret où le nombre de valeurs est pair (moyenne arithmétique des deux valeurs centrales)

9. Pour obtenir la boîte à moustaches, changer le nom du diagramme. Il peut aussi être transféré dans la fenêtre graphique pour plus d'options sur le quadrillage, les unités sur les axes, etc.



9. Utilisation d'un tableur dans le cas continu (GeoGebra)



Énoncé

Voici le tableau que l'on peut télécharger sur le site de l'Institut National des Statistiques. Il donne les chiffres de la population masculine en 2018, suivant des classes de 5 ans.

Tranche d'âge	Nombre d'hommes au 1/1/2018	Tranche d'âge	Nombre d'hommes au 1/1/2018
[0,5[317941	[50,55[410459
[5,10[341164	[55,60[390533
[10,15[327244	[60,65[343160
[15,20[321181	[65,70[296756
[20,25[339926	[70,75[239340
[25,30[371335	[75,80[164260
[30,35[362599	[80,85[129004
[35,40[375548	[85,90[74643
[40,45[367856	[90,95[24825
[45,50[395620	[95,100[4314
		Plus de 100	198

- Construire l'histogramme. Calculer la classe modale, la moyenne d'âge, l'écart-type.
- Construire le polygone des effectifs cumulés et calculer la valeur de la médiane et des quartiles.
- Calculer le pourcentage d'hommes s'écartant de la moyenne de plus de 2 fois l'écart-type.

Méthodologie :

Pour traiter ce genre de problème avec des données réelles assez grandes, il est préférable d'utiliser un tableur.

- Calculer la classe modale, la moyenne et l'écart-type dans ce cas permet de construire la formule de calcul de ces indices dans le cas continu. Prendre le centre de classes comme valeur représentative de la classe rejoint l'idée que la moyenne représente le fait que toutes les valeurs de la classe peuvent être assimilées à cette valeur.
- Pour construire le polygone des effectifs cumulés, il faut construire le tableau groupé en ajoutant la colonne des effectifs cumulés. En réalisant un ajustement, on peut calculer la médiane.
- Le calcul des quartiles se fait également sur le polygone. Les valeurs obtenues permettent de construire la boîte à moustaches.
- Répondre à cette question sur le polygone des effectifs cumulés puis introduire le théorème de Tchebychev pour montrer sa rapidité d'approximation lorsque l'on ne construit pas le polygone.

Construction de l'histogramme, calcul de la moyenne et de l'écart-type.

1. Fermer la fenêtre algèbre et ouvrir la fenêtre Tableur (Menu Affichage : Cocher Tableur et décocher Algèbre).
2. Entrer les valeurs des débuts de classe dans la colonne A et les valeurs de fin de classe dans la colonne B. Dans la colonne C, encoder les effectifs.
3. Sélectionner la première colonne puis cliquer sur la deuxième icône (l'icône bleue des graphiques) et sélectionner « Statistiques à une variable ».
4. Une nouvelle fenêtre apparaît. Cette petite fenêtre comporte une roue dentée à droite pour choisir le type de données. Choisir ici « Classes et effectifs ».
5. Cliquer sur la petite main de la colonne de droite. Apparaît alors dans la colonne de gauche des classes. Régler ensuite le début (30) et la largeur des classes (20). Appuyer sur enter.

N.B. Il est possible que GGB ne permette pas directement de changer la largeur des classes (comme dans l'exemple). Si c'est le cas, on peut le faire plus tard.

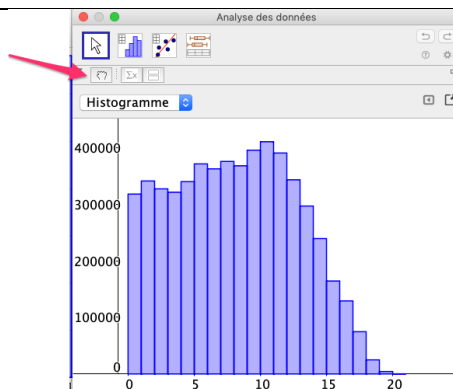


	A	B	C
1	0	5	317941
2	5	10	341164
3	10	15	327244
4	15	20	321181
5	20	25	339926
6	25	30	371335
7	30	35	362599
8	35	40	375548
9	40	45	367856
10	45	50	395620
11	50	55	410459
12	55	60	390533
13	60	65	343160
14	65	70	296756
15	70	75	239340
16	75	80	164260
17	80	85	129004
18	85	90	74643
19	90	95	24825
20	95	100	4314
21	100	105	198

	A	B	C	D	E	F
1	0	5				
2	5	10				
3	10	15				
4	15	20				
5	20	25				
6	25	30				
7	30	35				
8	35	40				
9	40	45				
10	45	50				
11	50	55				
12	55	60				
13	60	65				
14	65	70				
15	70	75				
16	75	80				
17	80	85	129004			
18	85	90	74643			
19	90	95	24825			
20	95	100	4314			
21	100	105	198			

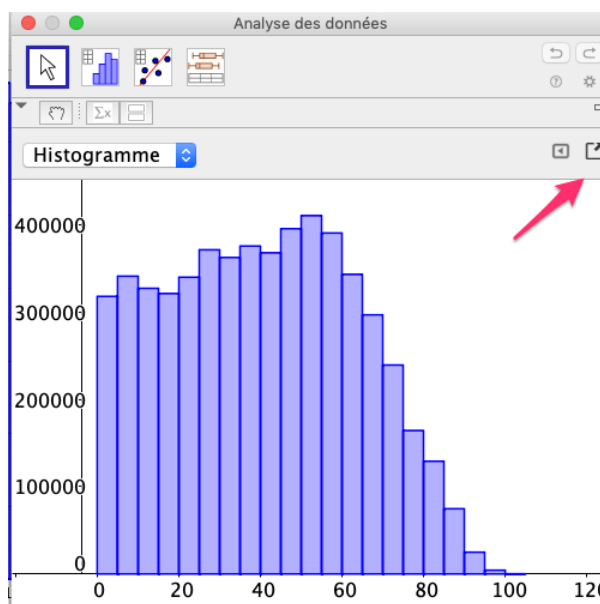
	A	B	C	D	E	F
1	0	5				
2	5	10				
3	10	15				
4	15	20				
5	20	25				
6	25	30				
7	30	35				
8	35	40				
9	40	45				
10	45	50				
11	50	55				
12	55	60				
13	60	65				
14	65	70				
15	70	75				
16	75	80				
17	80	85	129004			
18	85	90	74643			
19	90	95	24825			
20	95	100	4314			
21	100	105	198			
22						


6. Sélectionner la colonne des effectifs puis cliquer sur la petite main (à côté du mot Effectifs), les valeurs des effectifs apparaissent alors dans la deuxième colonne.

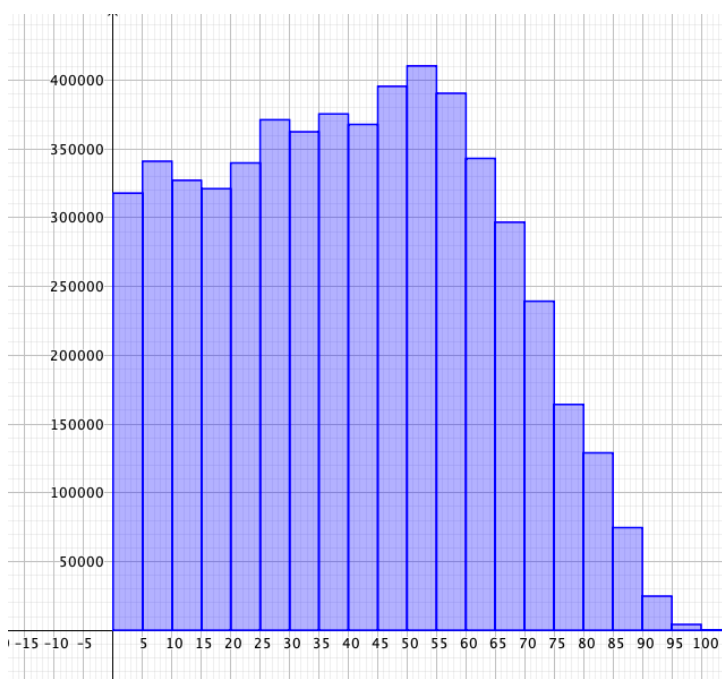


7. Cliquer sur Analyse pour obtenir l'histogramme.

N.B. Si les classes ne sont pas bien réglées (comme dans l'exemple), changer ici en cliquant de nouveau sur la petite main et en réglant la largeur à 5. On obtient alors l'histogramme correctement dessiné.



8. Copier l'histogramme vers la fenêtre graphique. (appuyer sur la touche ).



9. Régler avec la molette de la souris pour obtenir un graphique visible.
Si nécessaire, changer les valeurs des axes (clic droit sur les axes ou Options-Avancé-Graphique)

10. Pour obtenir les valeurs de la moyenne, de l'écart-type, il faut travailler comme dans le cas discret et réintroduire les centres des classes (colonne D)

D
2.5
7.5
12.5
17.5
22.5
27.5
32.5
37.5
42.5
47.5
52.5
57.5
62.5
67.5
72.5
77.5
82.5
87.5
92.5
97.5
102.5

11. Sélectionner « Statistiques à une variable » puis Classes et effectifs, ainsi que les deux colonnes l'une à la suite de l'autre avec les petites mains.

C	D	E	F	G	H	I
317941	2.5					
341164	7.5					
327244	12.5					
321181	17.5					
339926	22.5					
371335	27.5					
362599	32.5					
375548	37.5					
367856	42.5					
395620	47.5					
410459	52.5					
390533	57.5					
343160	62.5					
296756	67.5					
239340	72.5					
164260	77.5					
129004	82.5					
74643	87.5					
24825	92.5					
4314	97.5					
198	102.5					

12. Cliquer sur Analyse pour obtenir les valeurs calculées par GeoGebra.

n	55979...
Moyenne	40.4964
σ	23.0576
s	23.0576
Σx	22669...
Σx^2	12156...
Min	2.5
Q1	22.5
Médiane	42.5
Q3	57.5
Max	102.5

Attention !!

Les valeurs données par GGB pour la médiane et les quartiles ne sont pas correctes, car dans ce cas, le logiciel considère les valeurs comme étant discrètes, alors qu'elles sont continues. Il faut travailler avec le polygone des effectifs cumulés pour obtenir des valeurs correctes.

N.B. On peut obtenir les valeurs de la moyenne et de l'écart-type en créant des listes (sélectionner la colonne, clic-droit, créer liste) puis en tapant dans la ligne de saisie : Moyenne(L_1,L_2), Ecart-type(L_1,L_2).

Construction du polygone des effectifs cumulés, calcul de la médiane et des quartiles.

1. Pour tracer le polygone des effectifs cumulés croissants, il faut prendre pour valeur de référence l'extrémité de chaque classe. Il faut donc placer une ligne supplémentaire.
2. Pour cela, insérer une ligne en se plaçant sur la première ligne puis avec un clic droit, insérer une ligne.
3. Placer ensuite la valeur de début du premier intervalle (ici, 0) dans la ligne B1.
4. Construire ensuite la colonne des effectifs cumulés. Pour cela, dans la ligne 2, insérer une formule en commençant par =, qui va additionner les valeurs de D1 et C2. Tirer ensuite sur le bord extérieur droit de la cellule vers le bas jusqu'à la dernière ligne.
5. Construire dans la colonne E, les points qui apparaîtront sur le graphique. Pour cela, insérer la formule dans la cellule E1 $= (B1, D1)$. Puis tirer jusqu'à la dernière ligne. GeoGebra a construit des points qui vont directement apparaître dans la fenêtre graphique.

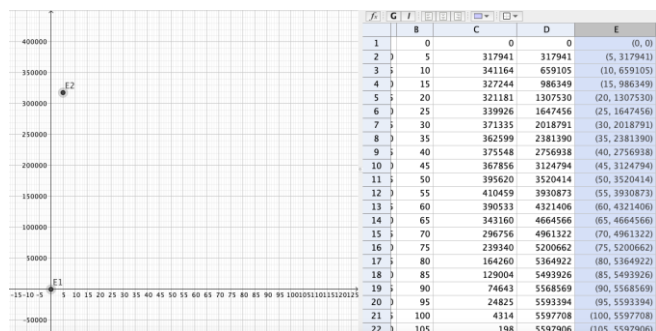
	A	B	C
	0	5	317941
	5	10	341164
	10	15	327244
	15	20	321181
	20	25	339926
	25	30	371335
	30	35	362599
	35	40	375548

	A	B	C	D
1		0		0
2	0	5	317941	$= D1 + C2$
3	5	10	341164	

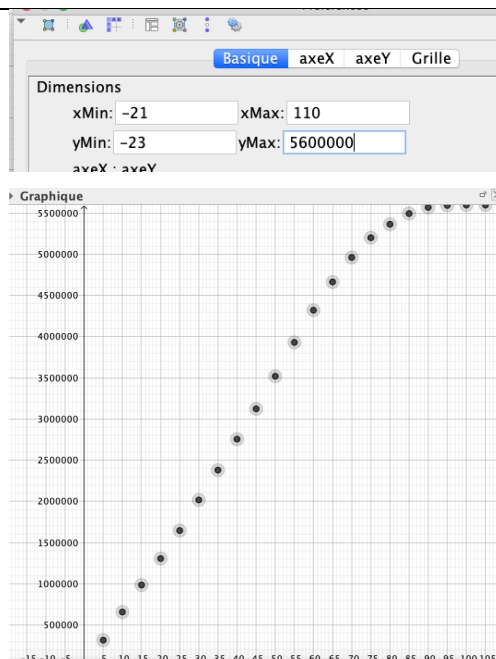
	A	B	C	D
1				
2	0	5	317941	317941
3	5	10	341164	659105
4	10	15	327244	986349
5	15	20	321181	1307530
6	20	25	339926	1647456
7	25	30	371335	2018791
8	30	35	362599	2381390
9	35	40	375548	2756938

	B	C	D	E
	0		0	$= (B1, D1)$
	5	317941	317941	

	B	C	D	E
1	0		0	(0, 0)
2	5	317941	317941	(5, 3179...
3	10	341164	659105	(10, 659...
4	15	327244	986349	(15, 986...
5	20	321181	1307530	(20, 130...
6	25	339926	1647456	(25, 164...
7	30	371335	2018791	(30, 201...
8	35	362599	2381390	(35, 238...
9	40	375548	2756938	(40, 275...

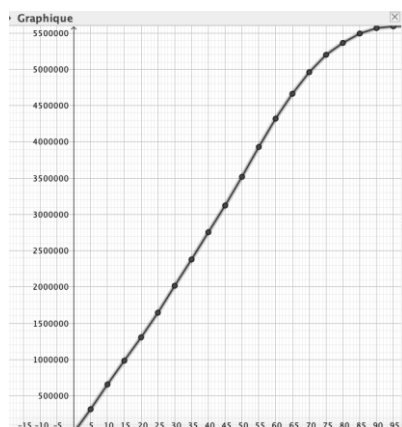


6. Pour les voir, ne pas oublier de régler les valeurs des deux axes. (clic droit sur les axes ou Affichage, Aspect, Préférences Graphe).



	B	C	D	E	F
1	0	0	0	(0, 0)	
2	5	317941	317941	(5, 3...	segment[E1,E2

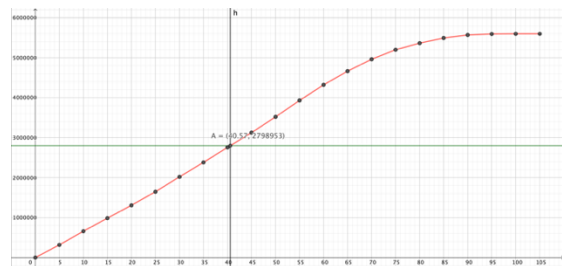
7. Pour ne pas voir apparaître le nom des points, sélectionner la colonne, puis clic droit et enlever le « v » à côté de « Afficher l'objet ». La fenêtre graphique est alors composée de tous les points qu'il faut alors relier.



8. Créer des segments reliant les différents points. Pour cela, dans la cellule F1, indiquer =segment puis cliquer consécutivement sur les cellules E1 puis E2. Tirer ensuite la formule vers le bas. Les segments se forment donc le polygone.

9. Pour approximer la médiane, on peut faire tracer une droite horizontale ayant comme équation $y = 2798953$ (la moitié de l'effectif total), en entrant cette équation dans le champ de saisie en bas de l'écran. On voit que la médiane vaut plus ou moins 41 ans. Pour obtenir la valeur exacte, on cherche l'intersection entre cette droite et le segment correspondant de la ligne brisée. On obtient alors la valeur exacte de la médiane.

Saisie: $y=2798953$



10. On obtient de manière semblable les quartiles sur le polygone des effectifs cumulés.

11. Pour répondre à la question c) on peut tracer deux droites verticales aux valeurs extrêmes $\bar{x} - 2\sigma, \bar{x} + 2\sigma$ et calculer le pourcentage d'hommes se trouvant à l'intérieur de cet intervalle d'âges.

On obtient 98,57% des hommes dans cet intervalle

