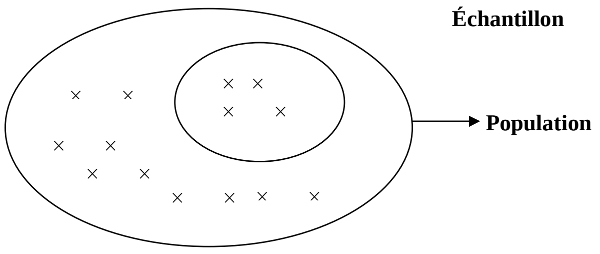
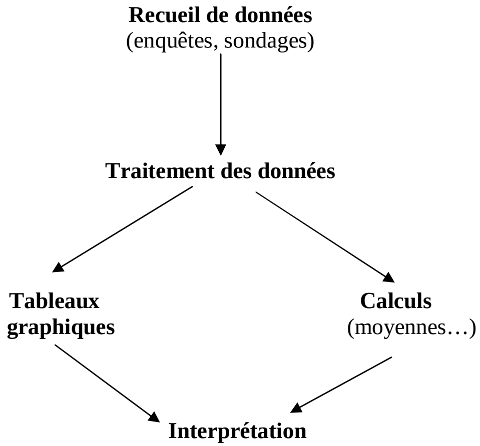
1ère S

Statistiques

I. Présentation générale



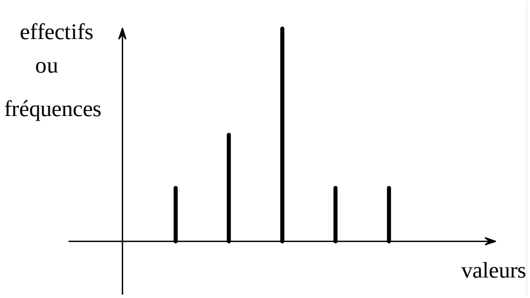


II. Vocabulaire usuel des statistiques

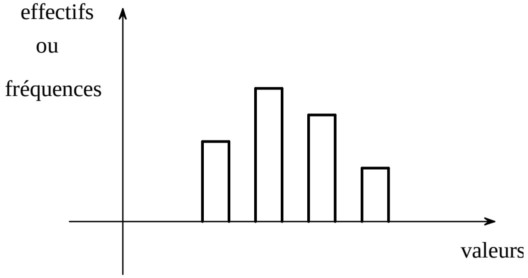
Notion	Exemples
Population : ensemble sur lequel porte l'étude	Etude de la répartition des cylindrées pour les voitures en France Etude sur l'ensemble des voitures en France
Individus : éléments qui composent la population	Les voitures
Caractère étudié : aspect que l'on observe pour les individus	Les cylindrées
Echantillon : partie de la population Il faut prendre un **échantillon représentatif** sinon les résultats seront faussés	Il faudrait prendre les voitures de toute la France
Série statistique : ensemble des valeurs collectées	
Série qualitative : quand le caractère n'est pas mesurable par des nombres	
Série quantitative : quand le caractère étudié fait l'objet d'une mesure, est mesurable par un nombre. On parle alors plutôt de **variable quantitative.**	Etude des pointures
Une **variable quantitative** peut être - **discrète** c'est-à-dire qu'elle ne peut prendre que des valeurs isolées ; - **continue** c'est-à-dire que le caractère étudié peut prendre toutes les valeurs dans un intervalle.	Nombre d'enfants d'une famille Température en France
Effectif total : nombre total d'individus dans la population	
Effectif d'une valeur : nombre d'individus qui présentent cette valeur. Parfois, on regroupe les valeurs en **classes** ; c'est le cas lorsque l'on a une variable quantitative continue.	
Fréquence d'une valeur : quotient de l'effectif de la valeur sur l'effectif total.	
Si une série statistique est connue par ses fréquences, on parle de **distribution de fréquences**.	

III. Présentation des résultats en graphiques (pour mémoire)

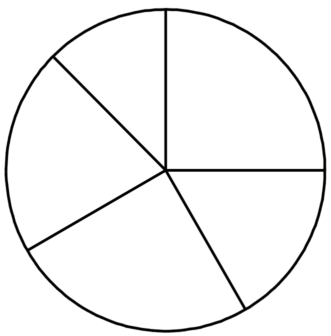
1°) Diagrammes en bâtons



2°) Diagrammes en barres



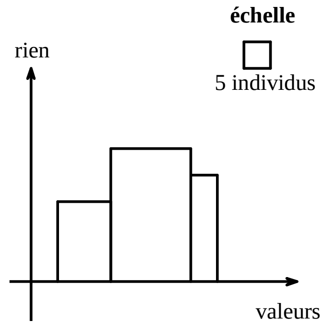
3°) Diagrammes circulaires ou semi-circulaires (« camemberts »)



Méthode pour déterminer les angles par rapport aux effectifs ou aux fréquences :
On sait que l'effectif total ou 100 % correspond à la totalité du cercle c'est-à-dire 360 °.

4°) Histogrammes

Série statistique dont les données sont regroupées en classes.
Représentation graphique par des rectangles accolés dont **les aires sont proportionnelles aux effectifs**.



2 cas

Histogramme à « pas » constant
(toutes les classes ont la même amplitude)

Histogramme à « pas » non constant

Utilisation d'un tableur très utile (permet de visualiser les résultats très rapidement).

On verra aussi les diagrammes en boîtes.

IV. Définitions des paramètres d'une série statistique (lexique)

Valeurs	x_1	x_2	…	x_p	
Effectifs	n_1	n_2	…	n_p	N

↓ **Effectif total**

N.B. :
- toujours les **valeurs** sur la **1ère ligne (ce sont de nombres quelconques)**
- toujours les **effectifs** sur la **2e ligne (ce sont des <u>entiers naturels</u>)**

Fréquence en %	$\dfrac{\text{effectif de la valeur}}{\text{effectif total (N)}}\times 100$
Moyenne	$\overline{x}=\dfrac{n_1x_1+n_2x_2+...+n_px_p}{\text{N}}$ Lorsque l'on ne tient pas compte des valeurs extrêmes ou aberrantes, on parle de **« moyenne élaguée »**.
Médiane	**Un** nombre tel qu'au moins 50 % des valeurs soient inférieures ou égales à ce nombre et 50 % des valeurs soient supérieures ou égales à ce nombre.
Étendue	Valeur maximale $-$valeur minimale
Mode(s)	La (ou les) valeur(s) de la série de plus grand effectif Il peut y en avoir plusieurs ; on parle alors de série **« unimodale »** ou **« multimodale »**
Quartiles	□ **1er quartile** : **la** plus petite valeur de la série telle qu'au moins 25 % des valeurs soient inférieures ou égales à cette valeur □ **3e quartile** : **la** plus petite valeur de la série telle qu'au moins 75 % des valeurs soient inférieures ou égales à cette valeur
Intervalle interquartile	$[Q_1\ ;\ Q_3]$
Écart interquartile	$\text{I}=Q_3-Q_1$
Variance	**Formule de définition** $\text{V}=\dfrac{n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+...+n_p\left(x_p-\overline{x}\right)^2}{\text{N}}$ On a donc $\text{V}\geqslant 0$. **Formule de calcul (propriété de Kœnig-Huyghens)** $\text{V}=\dfrac{n_1\left(x_1\right)^2+n_2\left(x_2\right)^2+...+n_p\left(x_p\right)^2}{\text{N}}-\left(\overline{x}\right)^2$
Écart-type	$\sigma=\sqrt{\text{V}}$ (σ : sigma minuscule)

V. Détermination de la médiane (cas d'une série quantitative discrète)

1°) Principe

□ Il faut toujours ranger les valeurs <u>dans l'ordre croissant</u>.
□ Le principe de la médiane est de séparer en 2 groupes de même effectif.

« Il faut acquérir la logique des statistiques ».

Exemple :

x_i	5,3	7,1	7,4	
n_i	4	3	2	N = 9

5,3 ; 5,3 ; 5,3 ; 5,3 ; 7,1 ; 7,1 ; 7,1 ; 7,4 ; 7,4

La médiane est 7,1.
<u>Donner du sens à la médiane</u>
<u>Interpréter</u>

Au moins 50 % des valeurs sont supérieures ou égales à 7,1.
Au moins 50 % des valeurs sont inférieures ou égales à 7,1.

- En géométrie, on appelle médiane d'un triangle une droite passant par un sommet et le milieu du côté opposé.
- On démontre aisément qu'une médiane partage un triangle en deux triangles de même aire.
- Cela montre l'analogie entre la médiane en statistiques et en géométrie.

2°) Exemple

Valeurs rangées dans l'**ordre croissant**

2 cas

Effectif total pair

<u>Exemple</u>

$N = 24$

$24 = 2 \times 12$

12 ; 12

$\text{Med} = \dfrac{12^e \text{ valeur} + 13^e \text{ valeur}}{2}$

Effectif total impair

<u>Exemple</u>

$N = 23$

$23 = 2 \times 11 + 1$

11 ; 1 ; 11

$\text{Med} = 12^e \text{ val}$

Il ne s'agit à priori pas d'une valeur de la série.

3°) Conventions

Effectif total pair : la médiane tombe entre 2 valeurs dont on fait la moyenne.

Effectif total impair : la médiane est la valeur du milieu.

4°) Méthode pratique

Utilisation des effectifs cumulés croissants

Exemple

Valeurs	61,3	61,4	61,5	61,6	
Effectifs	2	7	7	1	**Effectif total = 17**
Effectifs cumulés croissants	2	9	16	17	
Rang des valeurs	$1^{ère}$ et 2^e valeur	3^e à 9^e valeur	10^e à 16^e valeur	17^e valeur	

① $\text{Effectif total} = 17$
Nombre impair

② $17 = 2 \times 8 + 1$

8 ; 1 ; 8

③ $\text{Med} = 9^e \text{ valeur}$

61,3 ; 61,3 ; 61,4 ; 61,4 ; 61,4 ; 61,4 ; 61,4 ; 61,4 ; 61,4….

On remonte à la première ligne pour voir quelle est la 9^e valeur.

Med = 61,4

5°) Remarque

La médiane est toujours dans la même unité que les valeurs.

Par exemple, s'il s'agit de tailles en cm, la médiane sera en cm.

À titre informatif, le revenu médian en France, stable depuis plusieurs années, est de 1500 € net par mois.

VI. Détermination des quartiles (cas d'une série quantitative discrète)

1°) Principe

On cherche à partager le groupe en 4.

$$\begin{array}{ccc} | & || & | \\ Q_1 & \text{Med} & Q_3 \end{array}$$

(**Par convention**, le 2^{e} quartile est la médiane).

Le vendredi 8 février 2019

En langage naturel, on parle du 1er quartile comme étant l'intervalle.
« Je suis dans le premier quartile ». Ce n'est pas ce que l'on entend en mathématiques.

Les quartiles et la médiane déterminent 4 intervalles $[\text{Min} ; Q_1]$, $[Q_1 ; \text{Me}]$, $[\text{Me} ; Q_3]$, $[Q_3 ; \text{Max}]$.

2°) Exemple

Valeurs	30	45	50	60	61	
Effectifs	2	3	2	2	2	**Total = 11**
Effectifs cumulés croissants	2	5	7	9	11	

Médiane

① Effectif total $N = 11$
Nombre impair

② $11 = 5 \times 2 + 1$
$5 ; 1 ; 5$

③ $\text{Med} = 6^{e}$ valeur
$\text{Med} = 50$

Quartiles

1er quartile :

$\frac{N}{4} = \frac{11}{4} = 2{,}75$

La méthode consiste à toujours prendre l'entier immédiatement supérieur quand le résultat ne tombe pas juste.

$Q_1 = 3^{e}$ valeur $= 45$ (on remonte à la 1ère ligne)

3^{e} quartile :

$\frac{3N}{4} = 3 \times \frac{11}{4} = 8{,}25$

$Q_3 = 9^{e}$ valeur $= 60$

Interpréter les valeurs des quartiles :

D'après la définition du premier quartile, 45 est **la** plus petite valeur de la série telle qu'au moins 25 % des valeurs soient inférieures ou égales à cette valeur.

D'après la définition du troisième quartile, 60 est **la** plus petite valeur de la série telle qu'au moins 75 % des valeurs soient inférieures ou égales à cette valeur.

(Comparer les articles dans la définition de « la » médiane et des quartiles d'une série statistique).

Écart interquartile

$I = Q_3 - Q_1 = 60 - 45 = 15$

3°) Méthode

Attention : pas la même règle que pour « la » médiane

1er quartile

On calcule $\frac{N}{4}$.

$\rightarrow$ rang du quartile

3^{e} quartile

On calcule $\frac{3N}{4}$.

$\rightarrow$ rang du quartile

Quand le résultat ne tombe pas juste, on prend toujours l'entier immédiatement supérieur.

On regarde ensuite les valeurs correspondantes sur la 1ère ligne grâce aux effectifs cumulés croissants.

4°) Bêtise à ne pas faire

Ne pas écrire $Q_1 = \frac{N}{4}$ ou $Q_3 = \frac{3N}{4}$.

5°) Remarque

Les quartiles sont toujours dans la même unité que les valeurs.

Environ la moitié des valeurs appartiennent à l'intervalle interquartile.

6°) Déciles, centiles (quantiles)

Même principe de calcul que pour les quartiles.

□ **déciles**

On note D_1 le premier décile, D_2 le deuxième décile, D_3 le troisième décile, …, D_9 neuvième décile.

On les obtient en calculant $\left(\frac{\text{N}}{10}, \frac{2\text{N}}{10}, ...\right)$

On définit de même **l'intervalle interdécile** $\left[D_1 \,;\, D_9\right]$ et **l'écart interdécile** $D_9 - D_1$.

Au moins 80 % des valeurs appartiennent à l'intervalle interdécile.

□ **centiles**

C_1, C_2, C_3, …, C_{99} $\left(\frac{\text{N}}{100}, \frac{2\text{N}}{100}, ...\right)$

On définit de même **l'intervalle intercentile** $\left[C_1 \,;\, C_{99}\right]$ et **l'écart intercentile** $C_{99} - C_1$.

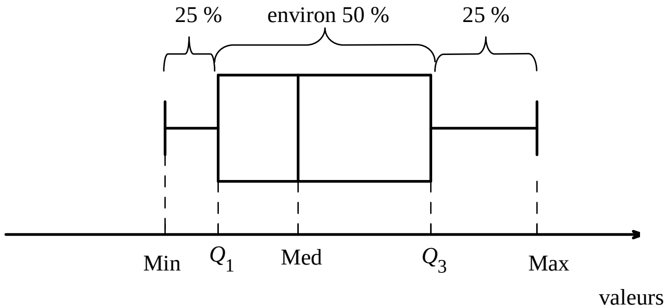
Au moins 90 % des valeurs appartiennent à l'intervalle intercentile.

VII. Diagrammes en boîte ou à moustaches

John Wilder Tukey (1915-2000) est un statisticien américain qui a ouvert la voie de l'analyse exploratoire de données. Il a créé les boîtes de dispersion (*box plot*) en 1977.

1°) Principe

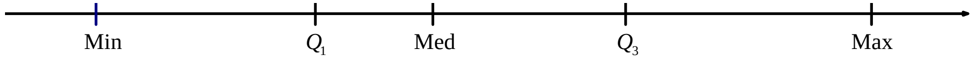
$$\left.\begin{array}{l}\text{Min}\\ Q_1\\ \text{Med}\\ Q_3\\ \text{Max}\end{array}\right\}\ \text{5 paramètres}$$



Il est important de comprendre qu'un diagramme en boîte donne une image résumée, un mode de représentation de la série statistique (répartition des effectifs / valeurs).

On rappelle l'importance de faire la différence entre valeurs et effectifs.

2°) Méthode



- On trace un axe d'échelle, horizontal ou vertical. On choisit une unité.
- Sur cet axe, on place les valeurs dans l'ordre croissant, à savoir : minimum, Q_1, médiane, Q_3, maximum.
- On trace une « boîte » rectangulaire allant de Q_1 à Q_3 séparée par une cloison au niveau de la médiane. La hauteur de la boîte est laissée au choix.
- Enfin, on trace les deux « moustaches » qui relient au maximum et au minimum. On les arrête par deux petits traits verticaux.
- Pour une meilleure lisibilité, tracer en pointillés des traits qui relient aux valeurs de l'axe d'échelle.

3°) Propriété

En général, il y a environ la moitié des valeurs dans la boîte (sauf cas exceptionnels de boîtes sans moustaches, par exemple).

4°) Intérêt

Comparaison à vue de deux (ou plus) séries statistiques en traçant les diagrammes en boîte l'un en dessous de l'autre avec un même axe d'échelle pour les deux.
(Voir exercices)

5°) Attention

- Mettre la légende sur l'axe d'échelle.
- Mettre la flèche sur l'axe d'échelle.
- Ne pas se tromper sur les valeurs (Q_1 et Q_3 sur la première ligne).

6°) Retour sur l'intervalle interquartile

• Propriété :

- Au moins la moitié des valeurs appartiennent à l'intervalle $\left[Q_1 \,;\, Q_3\right]$.
- Au moins la moitié des valeurs n'appartient pas à l'intervalle $\left]Q_1 \,;\, Q_3\right[$.

• Remarque :

La proportion de valeurs appartenant à l'intervalle $\left[Q_1 \,;\, Q_3\right]$ peut être assez éloignée de 50 %. Il peut même se faire qu'elle soit égale à 100 % lorsque le premier quartile et le troisième quartile sont respectivement égaux au minimum et au maximum de la série (par exemple dans le cas d'une série statistique où toutes les valeurs sont égales).

On démontre aisément le résultat suivant :

> Chacun des intervalles $[\text{Min} \,;\, Q_1]$, $[Q_1 \,;\, \text{Med}]$, $[\text{Med} \,;\, Q_3]$, $[Q_3 \,;\, \text{Max}]$ contient au moins le quart des valeurs.

VIII. Variance. Écart-type

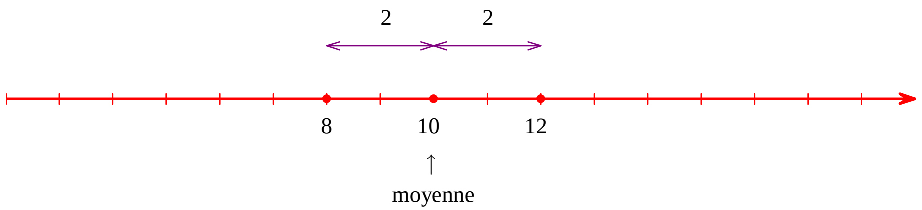
1°) Exemple introductif

Prenons l'exemple de deux élèves.
On regarde leurs notes sur un trimestre.

Élève 1 : 10 – 10 – 10
Élève 2 : 9 – 10 – 11
Les deux élèves ont tous les deux 10 de moyenne.

Néanmoins, le premier qui a eu toujours 10 est plus régulier que le second.

On peut visualiser les notes sur un axe.



On cherche à quantifier la dispersion des valeurs autour de la moyenne.

1ère idée :

On va calculer l'écart moyen de cet élève des deux élèves c'est-à-dire la moyenne des écarts à la moyenne de chacun des deux élèves.

On peut regarder le graphique qui est assez parlant.

On prend chaque valeur, on calcule l'écart par rapport à la moyenne. On calcule la moyenne.

$$\text{écart moyen de l'élève 1} = \frac{0+0+0}{3} = 0$$

$$\text{écart moyen de l'élève 2} = \frac{2+0+2}{3} = \frac{4}{3}$$

2e idée :

On va calculer la variance des notes des deux élèves c'est-à-dire la moyenne des écarts à la moyenne de chacun des deux élèves.

On prend chaque valeur, on calcule l'écart par rapport à la moyenne. On élève au carré. On calcule la moyenne.

$$V_1 = \frac{(10-10)^2+(10-10)^2+(10-10)^2}{3} = 0$$

$$V_2 = \frac{(8-10)^2+(10-10)^2+(12-10)^2}{3} = \frac{8}{3}$$

Les deux idées sont possibles mais c'est la deuxième idée que l'on retient pour mesurer la « dispersion » des valeurs autour de la moyenne. Ce choix n'est pas facile à justifier au lycée. Une justification apparaîtra cependant plus tard avec l'observation des moyennes d'échantillons.

2°) Définition

On considère une série statistique d'effectif total N avec des valeurs x_1, x_2, …, x_N (série statistique « brute »).

• On rappelle que la moyenne de la série est donnée par $\overline{x} = \dfrac{x_1+x_2+\ldots+x_N}{N} = \dfrac{\sum_{i=0}^{i=N} x_i}{N}$.

• La **variance** de la série est le nombre V défini par $V = \dfrac{\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\ldots+\left(x_N-\overline{x}\right)^2}{N} = \dfrac{\sum_{i=0}^{i=N}\left(x_i-\overline{x}\right)^2}{N}$.

C'est la moyenne des carrés des écarts des valeurs par rapport à la moyenne.
On peut dire que la formule de la variance est « parlante ».

• L'**écart-type** de la série statistique est le nombre noté σ égal à la racine carrée de la variance.
Autrement dit, $\sigma = \sqrt{V}$.

On peut démontrer la formule de Koenig-Huygens qui permet un calcul plus simple de la variance.

$$V = \frac{\sum_{i=0}^{i=n}\left(x_i\right)^2}{N} - \left(\overline{x}\right)^2$$

Cette formule exprime que la variance est égale à la moyenne des carrés des valeurs moins le carré de la moyenne.
Comme le nom l'indique, la variance permet de mesurer comment varient les valeurs autour de la moyenne.

3°) Formules dans le cas d'une série donnée par les valeurs et effectifs correspondant

Valeurs	x_1	x_2	…	x_p	
Effectifs	n_1	n_2	…	n_p	Total = N

Moyenne pondérée : $\overline{x} = \dfrac{n_1x_1+n_2x_2+\ldots+n_px_p}{N}$

Remarque :

La moyenne est notée $\overline{x}$ (x surmonté d'une barre).
Cette notation est une notation conventionnelle pour les moyennes en statistiques.
Cette notation n'a rien à voir avec une notation similaire utilisée en physique qui signifie valeur algébrique.

En utilisant la moyenne, on peut calculer la variance V et l'écart-type σ de **deux façons**.
Les deux formules donnent le même résultat.

□ **par la définition**

$$\boxed{\mathbf{V} = \frac{n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+\ldots+n_p\left(x_p-\overline{x}\right)^2}{\mathbf{N}}}$$

□ **par la formule de Kœnig-Huygens**

$\overline{x}$ désigne la moyenne dans les deux égalités

$$\boxed{\mathbf{V} = \frac{n_1\left(x_1\right)^2+n_2\left(x_2\right)^2+\ldots+n_p\left(x_p\right)^2}{\mathbf{N}}-\left(\overline{x}\right)^2}$$

Écart-type

$$\boxed{\sigma = \sqrt{\mathbf{V}}}$$

Commentaires :

● Pour calculer la variance (et donc aussi l'écart-type), il faut avoir calculé la moyenne auparavant.

● La variance sert à calculer l'écart-type.

● Pour calculer la variance, il vaut mieux utiliser la formule de Kœnig-Huygens.

● Par définition, la variance d'une série statistique est positive ou nulle. Cela permet donc de considérer sa racine carrée qui, par définition, est égale à l'écart-type.

● La variance (ou, ce qui revient au même, l'écart-type) d'une série statistique est nulle si et seulement si toutes les valeurs de la série sont égales.

4°) Exemple

Valeurs	1	2	3	4	
Effectifs	5	7	1	8	Total N = 21

● **Moyenne :**

$$\begin{aligned}\overline{x} &= \frac{5+7\times2+3+4\times8}{21}\\ &= \frac{54}{21}\\ &= \frac{18}{7}\end{aligned}$$

Avec la calculatrice, on obtient $\overline{x} = 2{,}57142857...$.
Donc $\overline{x} \approx 2{,}57$ (valeur arrondie au centième).

● **Variance :**

$$\begin{aligned}\mathrm{V} &= \frac{5\times1^2+7\times2^2+1\times3^2+8\times4^2}{21}-\left(\frac{18}{7}\right)^2 \text{ (formule de Kœnig-Huygens)}\\ &= \frac{170}{21}-\frac{324}{49}\\ &= \frac{7\times170-3\times324}{49\times3}\\ &= \frac{218}{147}\end{aligned}$$

Avec la calculatrice, on obtient $\mathrm{V} = 1{,}482993119...$. Donc $\mathrm{V} \approx 1{,}49$ (valeur arrondie au dixième).

Attention, on met la valeur au carré, pas l'effectif !

● **Écart-type :**

$$\begin{aligned}\sigma &= \sqrt{\mathrm{V}}\\ &= \sqrt{\frac{218}{147}}\end{aligned}$$

Avec la calculatrice, on obtient $\sigma = 1{,}217782081...$. Donc $\sigma \approx 1{,}22$ (valeur arrondie au dixième).

5°) Cas où les valeurs sont regroupées en classes

Pour calculer l'écart-type, on prend alors, comme pour la moyenne, le **centre de chaque classe**.

6°) Unités

L'écart-type est dans la **même unité que les valeurs** (comme la moyenne et la médiane).
Par exemple, s'il s'agit de tailles en cm, la moyenne, la médiane, les quartiles et l'écart-type seront aussi en cm.
La variance est dans l'unité au carré ; mais on ne le met pas car la variance ne sert qu'à calculer l'écart-type.

7°) À quoi servent la variance et l'écart-type ?

Donner la moyenne d'un contrôle est insuffisant. On devrait donner en plus pour avoir une bonne idée :
- les notes extrêmes (l'étendue) ;
- l'écart-type.

Exemple :
C'est souvent dans les matières scientifiques (maths-physique) que l'écart-type des notes est le plus grand contrairement aux matières littéraires où l'on a souvent un faible écart-type
En effet, souvent les notes en français sont concentrées autour de moyenne, entre 8 et 12 disons. Il est assez difficile d'avoir au-dessus de 14 ; il est également rare d'avoir des notes très faibles.
Alors qu'en mathématiques, on a souvent une très grande amplitude de notes.

Prenons l'exemple de deux élèves.
On regarde leurs notes sur un trimestre.

Élève 1 : 10 – 10 – 10
Élève 2 : 9 – 10 – 11
Les deux élèves ont tous les deux 10 de moyenne.

Néanmoins, le 1er qui a eu toujours 10 est plus régulier que le second.

L'écart-type des notes du 1er élève est égal à : $\sigma_1=\sqrt{\dfrac{(10-10)^2+(10-10)^2+(10-10)^2}{3}}=0$.

L'écart-type des notes du 2^{e} élève est égal à : $\sigma_2=\sqrt{\dfrac{(9-10)^2+(10-10)^2+(11-10)^2}{3}}=\sqrt{\dfrac{2}{3}}$.

On voit que l'écart-type des notes de l'élève 1 est 0 alors que l'écart-type des notes de l'élève 2 est $\sqrt{\dfrac{2}{3}}$.
Grâce à l'écart-type, on a pu mesurer le fait que l'élève 2 est moins régulier que l'élève 1.

8°) Propriété

L'écart-type d'une série statistique est toujours inférieur ou égal à l'étendue.

IX. Comparaisons des différents paramètres

1°) Couples de paramètres

On distingue 2 types de mesures :
paramètres

Mesures

de tendance centrale		**de dispersion**
moyenne	⟷	écart-type
médiane	⟷	écart interquartile (Q_3-Q_1) (étendue)

2°) Interprétation

• Plus l'écart-type est faible, plus les valeurs sont concentrées autour de la moyenne.

• Plus l'écart interquartile est faible, plus les valeurs sont concentrées autour de la médiane.

(Un écart-type nul signifie que toutes les valeurs sont égales).

3°) Application : comparaison de deux séries statistiques

Question d'élève (1ère S1) le vendredi 3 mars 2017

Qu'est-ce que comparer des séries statistiques ?

Méthode

On compare les mesures de tendance centrale puis les mesures de dispersion associées.

(On a besoin des 2 pour comparer).
La comparaison peut s'effectuer « graphiquement » grâce aux diagrammes en boîte (on compare alors les médianes et les écarts-types).

Voir exercices.

4°) Avantages-inconvénients

Le couple (médiane ; écart interquartile) est moins sensible que le couple (moyenne ; écart-type) aux variations des valeurs extrêmes.

X. Utilisation de la calculatrice

1°) Rentrer une liste

TI 82-83	**Casio**
$\boxed{\text{stats}}$ Edit	Menu 2 $\boxed{\text{STAT}}$

L1	L2	L3
List 1	List 2	List 3

On saisit les valeurs dans la 1ère colonne, les effectifs dans la 2^{e}.

Exemple :

Valeurs	1	2	3	4
Effectifs	5	7	1	8

On rentre les valeurs (nombres de la première ligne du tableau) dans la colonne L_1 et les effectifs (nombres de la deuxième ligne du tableau) dans la colonne L_2.

2°) Calcul des paramètres

$\boxed{\text{stats}}$ CALC

$\bar{x} = ...$ **moyenne**
$\sum x = ...$ on s'en fiche
$\sum x^2 = ...$ on s'en fiche
$S_x = ...$
$\sigma_x = ...$ **écart-type**

Attention, les conventions de calcul de la calculatrice ne sont pas les mêmes que celles données en cours pour la médiane et les quartiles. À enlever (le 12-3-2018)

$\boxed{\text{stats}}$
↓
$\boxed{\text{EDIT}}$ ⟶ « Edit »

Entrer les valeurs (liste L1 : valeurs x_i, liste L2 : effectifs n_i ; on n'utilise pas la liste L3)

Le mardi 9 mars 20187

Calculatrice TI 83 Premium CE
Il est possible de modifier la méthode de calcul des quartiles en appuyant sur $\boxed{\text{stats}}$ puis 6 .

Puis

$\boxed{\text{stats}}$
↓
$\boxed{\text{CALC}}$
↓
1- Var Stats
↓
Taper $\boxed{\text{L1}}$ $\boxed{,}$ $\boxed{\text{L2}}$ $\boxed{\text{entrer}}$

La calculatrice affiche :

Stats 1-Var

$\bar{x} = 2,571428571$
$\sum x = 54$
$\sum x^2 = 170$
$S_x = 1,247855303$
$\sigma_x = 1,217782081$
$n = 21$
$\min x = 1$
$Q_1 = 2$
$\text{Med} = 2$
$Q_3 = 4$
$\max x = 4$

Le mardi 9 mars 2018

Il faut changer l'exemple du cours. On ne voit pas bien la boîte car $Q_1 = \text{Med}$.

La calculatrice ne donne pas des valeurs exactes : elle donne des valeurs approchées. C'est le point négatif. Néanmoins, en général, en statistiques, les valeurs exactes n'ont pas d'intérêt.

La calculatrice ne donne pas la variance. Si on veut l'obtenir, on peut élever au carré le résultat de l'écart-type.

3°) Diagrammes

Il est possible d'obtenir différents diagrammes d'une série statistique.

• **Affichage du digramme en bâtons ou diagramme en barres**

Attention, la calculatrice accole les barres, ce qui fait ressembler le graphique à un histogramme ce qui n'est pas le cas.

• **Affichage du diagramme en boîte**

$\boxed{\text{stats}}$

1 : Edite (on rentre L1 et L2)

$\boxed{\text{2nde}}$ $\boxed{f(x)}$

1 : $\boxed{\text{entrer}}$

▷

Liste X : L1
Effectifs : L2

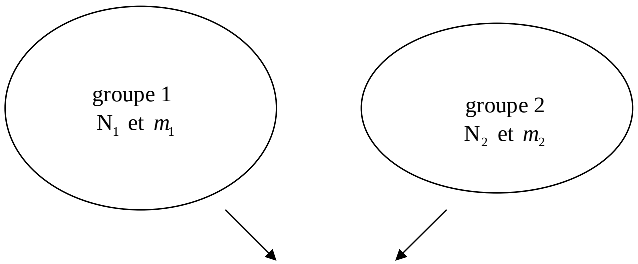
$\boxed{\text{trace}}$

XI. Calculs de la moyenne d'une série à l'aide des moyennes de deux sous-groupes

1°) Propriété

On considère une série statistique S_1 de moyenne m_1 et d'effectif total N_1.
On considère une série statistique S_2 de moyenne m_2 et d'effectif total N_2.
Alors la moyenne de la série statistique obtenue en réunissant les séries S_1 et S_2 est égale à

$$\boldsymbol{m = \frac{\mathrm{N}_1 \times m_1 + \mathrm{N}_2 \times m_2}{\mathrm{N}_1 + \mathrm{N}_2}}$$



$$\boldsymbol{m = \frac{\mathrm{N}_1 \times m_1 + \mathrm{N}_2 \times m_2}{\mathrm{N}_1 + \mathrm{N}_2}}$$

2°) Démonstration

Facile (voir cours de 2[e])

3°) Exemple

Un devoir

Une classe → 12 filles
Une classe → 8 garçons

Moyenne des filles $= 14$
Moyenne des garçons $= 16$

Moyenne de la classe $= \dfrac{4 \times 12 + 8 \times 16}{20} = 14{,}8$

4°) Remarque

Il n'existe pas de formule analogue pour les autres indicateurs tels que la variance, l'écart-type, la médiane etc.

XII. Formules de calculs de quelques indicateurs à l'aide des fréquences

1°) Moyenne

Valeurs	x_1	x_2	…	x_p	
Effectifs	n_1	n_2	…	n_p	Effectif total = N
Fréquences	$f_1 = \dfrac{n_1}{\mathrm{N}}$	$f_2 = \dfrac{n_2}{\mathrm{N}}$		$f_p = \dfrac{n_p}{\mathrm{N}}$	1

On a : $\boldsymbol{\overline{x} = f_1 \times x_1 + f_2 \times x_2 + \ldots + f_p \times x_p}$.

(La démonstration est très facile).

2°) Variance

Avec les mêmes notations qu'au 1°), on obtient les deux formules de la variance suivantes à partir de la définition et de la formule de Kœnig-Huygens.

$$\mathrm{V} = f_1 \times \left(x_1 - \overline{x}\right)^2 + f_2 \times \left(x_2 - \overline{x}\right)^2 + \ldots + f_p \times \left(x_p - \overline{x}\right)^2$$

$$\mathrm{V} = \left[f_1 \times \left(x_1\right)^2 + f_2 \times \left(x_2\right)^2 + \ldots + f_p \times \left(x_p\right)^2 \right] - \left(\overline{x}\right)^2$$

XIII. Détermination des indicateurs dans le cas d'une série de valeurs regroupées en classes

1°) Calcul de la moyenne et de la variance

On prend pour valeurs le centre de chaque classe.

2°) Détermination de la médiane et des quartiles

Méthode

On utilise une méthode graphique.
On utilise le **polygone des effectifs cumulés croissants** ou des fréquences cumulées croissantes.
Pour cela, on calcule d'abord les effectifs cumulés croissants (tableau) puis on trace le polygone des effectifs cumulés croissants : il s'agit d'un ligne brisée ouverte constituée de **segments de droites**.

● **Pour lire la médiane**, on place la valeur $\dfrac{\mathrm{N}}{2}$ sur l'axe des ordonnées ; on lit sur l'abscisse du point du polygone qui a pour ordonnée $\dfrac{\mathrm{N}}{2}$.

● **Pour lire le 1er quartile**, on place la valeur $\frac{N}{4}$ sur l'axe des ordonnées ; on lit sur l'abscisse du point du polygone qui a pour ordonnée $\frac{N}{4}$.

● **Pour lire le 3e quartile**, on place la valeur $\frac{3N}{4}$ sur l'axe des ordonnées ; on lit sur l'abscisse du point du polygone qui a pour ordonnée $\frac{3N}{4}$.

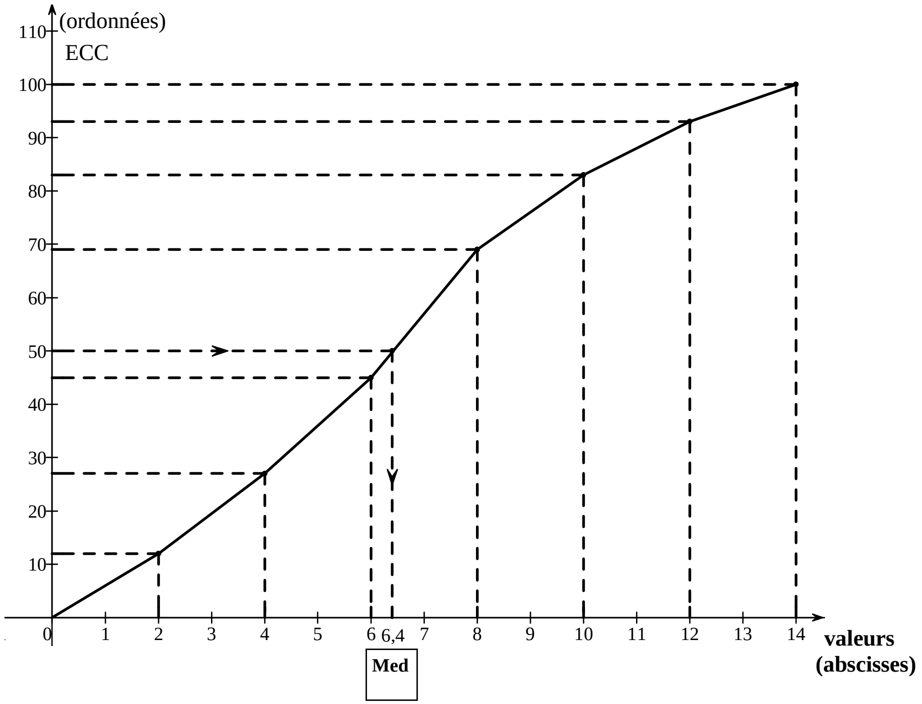
Cette méthode s'adapte évidemment sans problème aux quantiles (déciles, centiles …).

Exemple :

Le tableau ci-dessous donne la distribution des effectifs d'une série.

Classe	[0 ; 2[	[2 ; 4[	[4 ; 6[	|6 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 14[
Effectif	12	15	18	24	14	10	7
Effectif cumulé croissant	12	27	45	69	83	93	100
Borne supérieure	2	4	6	8	10	12	14

On détermine la médiane et les quartiles graphiquement grâce au polygone des effectifs cumulés croissants.



Détermination graphique d'une valeur approchée de la médiane.

On utilise le polygone des effectifs cumulés croissants.

L'effectif total est égal à 100.

On le divise par 2 : $\frac{100}{2} = 50$.

On se place à 50 d'ordonnée.
On va jusqu'au polygone.
On lit la valeur en abscisse.
On lit graphiquement $\text{Med} \approx 6,4$.

Détermination graphique d'une valeur approchée du 1er quartile.
On utilise le polygone des effectifs cumulés croissants.
L'effectif total est égal à 100.

On le divise par 4 : $\frac{100}{4} = 25$.
On se place à 25 d'ordonnée.
On va jusqu'au polygone.
On lit la valeur en abscisse.
On lit graphiquement $Q_1 \approx 3,5$.

Même méthode pour le 3e quartile.
On lit graphiquement : $Q_3 \approx 8,9$.
N.B. : On peut aussi calculer la médiane et les quartiles par interpolation linéaire (méthode d'interpolation linéaire qui sera vue en enseignement obligatoire).

4 documents complémentaires à lire impérativement :

- Intervalle interquartile et pourcentage d'appartenance

- Dossier Club Med

- Médiane et quartiles

- Texte sur les statistiques (copié par Jules de Senneville)

- Texte sur les statistiques issu du libre de seconde collection Déclic édition 2011 (Antoine de Parcieux)