UNITÉ 12: STATISTIQUE

DÉFINITION: La **statistique** est la branche des mathématiques qui a trait à la collecte, au classement, à l'analyse et à l'interprétation des données afin d'en tirer des conclusions et de faire des prévisions.

1. Population et échantillon. Caractères ou variables statistiques.

Recensement : recherche d'informations sur toute une population. Lorsque cette population est constituée d'objets, on parle plutôt d'**inventaire**.

Population : ensemble des êtres vivants ou des objets sur lesquels porte une étude statistique.
On se pose la question : «Sur qui porte l'étude? Quelle personne, quel animal ou quelle chose?».

Exemples de population :

- *Si on note l'âge des garçons de la classe, la population c'est les garçons.*
- *Si on relève leur couleur favorite, la population c'est aussi les garçons.*
- *Si j'étudie le sommet des montagnes, la population c'est les montagnes*
- *Si je note la vitesse des voitures, la population c'est les voitures.*

Échantillon : En statistique, un échantillon est un ensemble d'individus représentatifs d'une population.

Individu : En statistique, un individu est un élément d'un ensemble, généralement appelé « population », dont on mesure (ou observe) la valeur qu'il a pour la variable étudiée.

Une **série statistique** est simplement une liste de mesures obtenues généralement lors d'une étude ou de relevés de mesures. Pour avoir une série statistique, il suffit de produire une liste de nombres, ou de mesures.

Exemples de séries statistiques :

- *Une liste avec l'âge de chaque garçon forme une série statistique : 4, 12, 7, 8*
- *La liste des couleurs favorites des garçons, forme aussi une série statistique: rouge, rouge, vert, noire*
- *Avec la nationalité des garçons, on obtient une autre série statistique: française, marocaine, allemande, française*

Chaque nombre ou mesure d'une série statistique s'appelle **une valeur**.
Attention, une valeur peut être un nombre, mais ce n'est pas obligatoirement le cas.

Exemples de valeurs d'une série statistique :

- *Les nombres « 6 » et « 7 » sont des valeurs de la série: 4, 12, 7, 6, 8*
- *« rouge » et « vert » sont des valeurs de la série: rouge, rouge, vert, noire, rouge*
- *« française » , « allemande » sont des valeurs de la série: française, française, allemande, française, marocaine*

Les **valeurs extrêmes** sont les valeurs minimum et maximum dans la série.

Exemple :
Dans la série 1, 5, 8, 9, 12, les valeurs extrêmes sont 1 (le minimum) et 12 (la valeur maximum) .

Le caractère d'une série statistique c'est ce qui est mesuré ou étudié pour la population.
On se pose la question «Sur quoi porte l'étude ? Qu'est-ce qui est étudié ?»

Exemples de caractères :
- *Si je note l'âge des garçons, le caractère c'est l'âge.*
- *Si je relève leur couleur favorite, le caractère c'est leur couleur préférée.*
- *Si je mesure la hauteur des sommets de montagnes, le caractère c'est la hauteur.*
- *Si je mesure la vitesse des voitures, le caractère c'est la vitesse.*

Un caractère peut être :
- **qualitatif** si ses valeurs ne sont pas numériques. On appelle aussi **modalité.**
- **quantitatif** si ses valeurs sont numériques.

Exemples :
- *La taille est un caractère quantitatif, car c'est un nombre (une valeur numérique).*
- *La couleur préférée ou la nationalité sont des caractères qualitatifs car ce ne sont pas des nombres.*

Si la valeur du caractère ne peut prendre qu'une série de valeurs isolées, définies à l'avance, on dit que le caractère est **discret.**

Par contre, si la valeur du caractère peut prendre une infinité de valeur comprise entre 2 nombres, on dit que le caractère est **continu.** C'est-à-dire, on peut prendre toutes les valeurs d'un intervalle.

Exemples :
- *L'âge des adolescents peut prendre qu'une valeur parmi 13, 14, 15, 16, 17, 18. C'est un ensemble limité (isolé) de 6 valeurs. L'âge donc un caractère discret*

- *En revanche, la vitesse rigoureusement exacte d'une voiture comprise entre 0 et 230 km/h est un ensemble infini de valeurs qui peuvent être 20 km/h mais aussi 68,1 km/h mais encore 137,121354 km/h . La vitesse est un caractère continu.*

2. Effectif, fréquence et étendue. Tableau de fréquences.

Il faut distinguer :
L' effectif total : c'est le nombre de valeurs dans la série statistique. **N**
L' effectif d'une valeur donnée: c'est le nombre de fois ou la valeur apparaît pour cette série. f_i
L'effectif cumulé : Pour calculer un effectif cumulé, il suffit d'ajouter à l'effectif d'une valeur d'un caractère, le ou les effectifs des valeurs précédentes. $F_i = \sum_{j=1}^{i} f_j$
Remarque : On ne peut pas calculer l'effectif cumulé pour des variables qualitatives

Exemple 1:
Prenons la série 2, 2, 4, 34, 11, 4, 2
L'effectif total pour cette série est de 7 puisqu'il y a 7 valeurs (des nombres ici) .
L'effectif de la valeur 2 est de 3 puis que la valeur 2 apparaît 3 fois dans la liste.
Pour la valeur 4 c'est 2 , puisque le nombre 4 apparaît à 2 reprises.

Exemple 2:
Prenons la série: bleu, bleu, vert, noir, rouge, vert , bleu, noir, noir
L'effectif total pour cette série est de 9 puisqu'il y a 9 valeurs (Ici, des couleurs).
L'effectif de la valeur bleu est de 3 puisque le bleu apparaît 3 fois dans la liste.
Pour la valeur vert c'est 2 , puisque le vert apparaît à 2 reprises.

La fréquence d'une valeur donnée, c'est le quotient (la division) de l'effectif de la valeur par l'effectif total. Intuitivement, elle indique la proportion de la présence de la valeur dans la liste. $h_i = \frac{f_i}{N}$

Fréquence cumulée : Pour calculer une fréquence cumulée, il suffit d'ajouter à la fréquence d'une valeur d'un caractère, la ou les fréquences des valeurs précédentes.

$$H_i = \sum_{j=1}^{i} h_j = \sum_{j=1}^{i} \frac{f_j}{N}$$

Remarque : On ne peut pas calculer la fréquence cumulée pour des variables qualitatives

Exemple:
Prenons la série: 2, 2, 4, 34, 11, 4, 2, 1, 9, 9
L'effectif de la valeur 2 est de 3 puisque le nombre 2 apparaît 3 fois dans la liste.
L'effectif total pour cette série est de 10 puisqu'il y a 10 valeurs.
La fréquence = (effectif de la valeur / effectif total) = 3 / 10

L' étendue c'est la différence entre les valeurs extrêmes. Pour la calculer, il suffit de soustraire la plus grande valeur (maximum) et la plus petite valeur (minimum) de la série.

Exemple:
Prenons la série: 2, 2, 4, 34, 11, 4, 2, 1, 9, 9
Les valeurs extrêmes sont 11 (plus grande valeur) et 1 (la plus petite)
L'étendue pour cette série est de 11 – 1 =10 (Différence entre les valeurs extrêmes).

Remarque:
L'étendue donne des informations sur ce qu'on appelle la **dispersion** de la série statistique :
Si l'étendue est très petite, alors il y a peu d'écart entre toutes les valeurs de la série. Celle-ci est homogène.
Si au contraire l'étendue est grande, alors l'écart est important entre la plus petite et la plus grande valeur.

Tableau de fréquences

En statistique, on utilise souvent des **tableaux et des diagrammes** pour fournir des informations d'une manière claire et concise.

Comment établir le tableau d'une série statistique ?

Rappelez :
En statistique, on appelle population l'ensemble sur lequel on travaille.
Dans cette population, on étudie un caractère que l'on appelle variable statistique.
On étudie principalement des variables quantitatives, c'est-à-dire des variables qui prennent des valeurs numériques.

Dans le tableaux, les valeurs (caractère quantitatif) sont généralement présentées en ordre croissant; tandis que les modalités (caractère qualitatif) sont généralement présentées selon un certain ordre : alphabétique, chronologique ou d'effectif.

La variable quantitative peut être :
– soit discrète, quand elle prend un nombre fini de valeurs ;
– soit continue, quand elle prend toute valeur comprise entre deux nombres donnés.

Quand la variable statistique X est discrète, on compte, pour chaque valeur de X, le nombre d'individus prenant cette valeur ; c'est l'effectif de la valeur.
On aboutit à un tableau du type :

Valeur de X	x_1	x_2	x_3	…	x_p
Effectif	f_1	f_2	f_3	…	f_p

Quand la variable statistique X est continue, on regroupe les valeurs en classes.

- Les classes sont des intervalles semi-ouverts $[x_i, x_{i+1})$.
- Leur amplitude est le nombre $x_{i+1} - x_i$
- Leur centre est le nombre $\frac{x_i + x_{i+1}}{2}$
- Pour calculer leur amplitude d'une classe, on utilise la formule : $\frac{Valeur\ \ \max imum - Valeur\ \ \min imum}{\sqrt{N}}$
- Rappelez qu'il faut approcher le résultat.

Pour chaque classe, on compte le nombre d'individus qui prennent une valeur supérieure ou égale à x_i et inférieure à x_{i+1} : c'est l'effectif de la classe.
On aboutit à un tableau du type :

Valeur de X	$[x_1,x_2)$	$[x_2,x_3)$	$[x_3, x_4)$	…	$[x_p, x_{p+1})$
Effectif	f_1	f_2	f_3	…	f_p

Remarque:
• Quand le nombre de valeurs prises par la variable statistique est trop grand, on traite la variable discrète comme une variable continue.
• Quand on regroupe les valeurs par classes, on essaye d'avoir des classes de même amplitude et pas trop nombreuses. Mais, souvent, les valeurs extrêmes posent problème, c'est pourquoi les premières ou dernières classes sont soit ouvertes, soit d'amplitude différente des autres classes.

Tableau de fréquences

Variable	Effectif f_i	Effectif cumulé F_i	Fréquence $h_i=\frac{fi}{N}$	Fréquence cumulée $H_i=\frac{Fi}{N}$	Pourcentage $\%= h_i \cdot 100$
Xi ou [xi, xi+1)	f_i	F_i	h_i	H_i	%
TOTAL	$N=\sum_i f_i$		$\sum_i h_i =1$		100

Exemple:

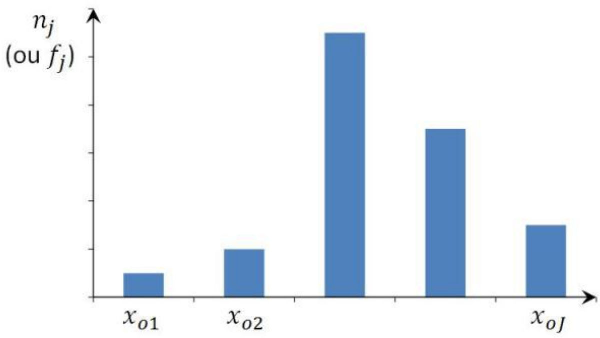
Résultats histoire-géographie dans une classe de 3ème

Classes	Centre de la classe	Effectifs f_i	Effectifs cumulés F_i	Fréquences h_i	Fréquences cumulées H_i	Pourcentage %
[0,2)	1	1	1	1/28	1/28	3,57%
[2,4)	3	3	4	3/28	4/28	10,71%
[4,6)	5	8	12	8/28	12/28=0,42	28,57%
[6,8)	7	12	24	12/28	24/28=0,85	42,86%
[8,10)	9	4	28	4/28	28/28=1	14,29%
Total		**N=28**		**1**		**100**

3. Diagrammes ou graphiques statistiques.

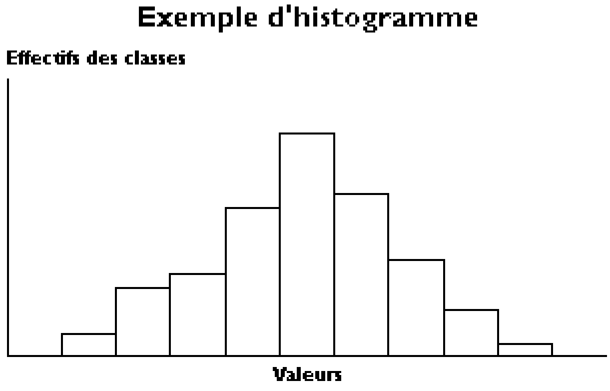
Diagramme en bâtons

Un diagramme en bâtons est une représentation graphique de données statistiques à l'aide de segments. Les valeurs du caractère étudié sont représentées sur l'axe horizontal, les effectifs sur l'axe vertical. À chaque valeur correspond un bâton. Les hauteurs des bâtons
sont proportionnelles aux effectifs représentés.



L'histogramme

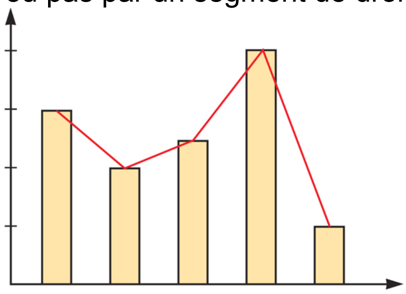
Pour représenter graphiquement une série statistique donnée par des classes, on utilise un histogramme, avec en abscisses les classes des valeurs du caractère. Un histogramme est constitué de rectangles juxtaposés ; la largeur de chaque rectangle correspond à l'intervalle de la classe correspondante ; sa hauteur est telle que l'aire du rectangle est proportionnelle à l'effectif de la classe.

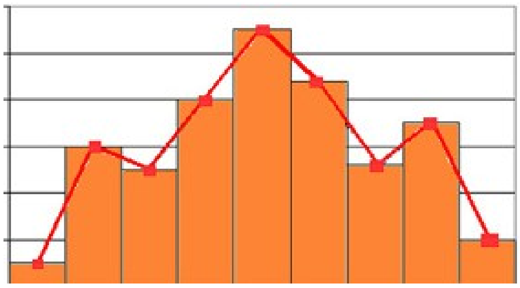


Le polygone des effectifs/fréquences

Le polygone de fréquences est une autre représentation graphique (en ligne brisée) de la distribution de fréquences d'une variable quantitative.

Pour tracer le polygone, on joint les points milieu du sommet des rectangles adjacents ou pas par un segment de droite.



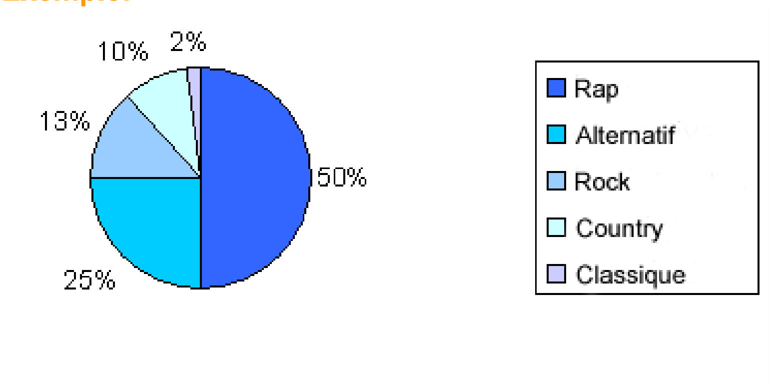


Le diagramme circulaire

Le diagramme circulaire donne pour chaque fréquence un secteur angulaire bien évidemment proportionnel. Après avoir déterminé les fréquences, on construit un tableau de proportionnalité entre les fréquences et les angles (sachant que 100% correspond à 360°).

L'angle s'obtient avec la formule $\mathbf{n^o = h_i \cdot 360^o}$

Exemple:



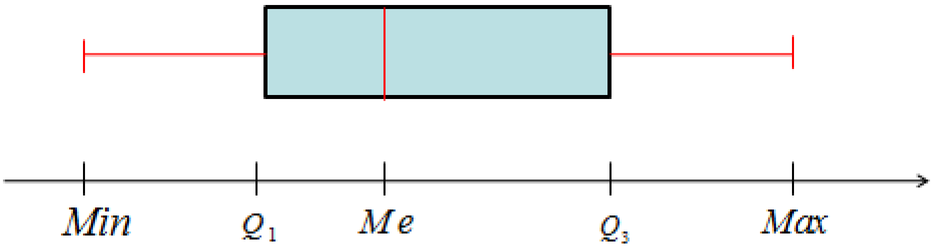
Remarque: Le **donut chart** est un camembert troué au milieu. Dans ce cas, c'est la longueur de l'arc de cercle correspondant à chaque catégorie qui représente la part de chaque catégorie dans le tout représenté.

Le choix de la représentation graphique statistique dépend de ce que l'on veut mettre en évidence:

- **Variable qualitative:** diagramme en bâtons ou diagramme circulaire.
- **Variable quantitative discrète**: diagramme en bâtons, polygone de fréquences ou diagramme circulaire.
- **Variable quantitative continue:** histogramme ou polygone de fréquences.

D'autres diagrammes :

Le **diagramme en boîte à moustaches** résume seulement quelques caractéristiques de position du caractère étudié (médiane, quartiles, Min/ Max ou déciles). Il est utilisé principalement pour comparer un même caractère dans deux populations de tailles différentes. Il s'agit de tracer un rectangle allant du premier quartile au troisième quartile et coupé par la médiane. On ajoute parfois des segments aux extrémités menant jusqu'aux valeurs min/max. On parle alors de diagramme en boîte à moustaches ou à pattes.



4. Les paramètres de tendance centrale

Mode: Le mode est la modalité ou la valeur qui a le plus grand effectif, la plus grande fréquence. Il est noté **Mo**.

Attention: La mode est la seule mesure de tendance centrale applicable aux variables qualitatives.

Remarque: Pour les variables quantitatives continues, on parle de classe modale, et le mode est leur centre. L'amplitude des classes doit être constante.

Exemple 1:
Perception de l'état de santé

X_i	Effectifs f_i
« très bon »	220
« bon »	345
« moyen »	155
« mauvais »	25
« très mauvais »	5
Total	**N=750**

La mode est la modalité *«bon»*.
L'effectif modal est 345

Exemple 2:
Résultats histoire-géographie dans une classe de 3ème

Classes	Effectifs f_i
[0,2)	1
[2,4)	3
[4,6)	8
[6,8)	12
[8,10)	4
Total	**N=28**

La classe modale est [6,8). Mo=7
L'effectif modal est 12

Moyenne : La moyenne est la mesure la plus commune de tendance centrale.
La moyenne est la somme des valeurs divisée par l'effectif total.
Il est noté $\overline{x}$

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i \cdot f_i}{N}$$

Avec :

x_i = valeur de la variable X (var. discrète) ou centre de la classe(var. continue)

f_i = effectif correspondant à la valeur xi

N = taille de l'échantillon (effectif total)

Exemple 1:

Nombre d'enfants par famille

X_i	Effectifs f_i	$X_i \cdot f_i$
0	3	0
1	8	8
2	12	24
3	5	15
4	2	8
Total	**N=30**	**55**

La moyenne est 1,83 par famille

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{N} = \frac{0 \cdot 3 + 1 \cdot 8 + 2 \cdot 12 + 3 \cdot 5 + 4 \cdot 2}{75} = \frac{55}{30} = 1{,}83$$

Exemple 2:

Résultats histoire-géographie dans une classe de 3ème

Classes	Centre de la classe	Effectifs f_i
[0,2)	1	1
[2,4)	3	3
[4,6)	5	8
[6,8)	7	12
[8,10)	9	4
Total		**N=28**

La moyenne est 6,07

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{N} = \frac{1 \cdot 1 + 3 \cdot 3 + 5 \cdot 8 + 7 \cdot 12 + 9 \cdot 4}{28} = \frac{170}{28} = 6{,}07$$

Médiane : La médiane est la valeur de la variable qui partage en 2 parties égales ou en 2 sous-ensemble égaux la population.
Elle divise une série statistique ordonnée en deux ensembles comportant chacun 50% des données. 50% des valeurs sont supérieures à la médiane et 50% inférieures.

Pour déterminer la médiane d'un échantillon ou d'une population (variable discrète) :
1.- on classe les individus par ordre croissant
2.- on prend celui du milieu
Avec N le nombre d'individus dans l'échantillon, la médiane porte le numéro d'ordre $\frac{N+1}{2}$ dans la suite des individus classés par ordre croissant.
Il y a deux possibilités:
* Si N est impair, la médiane est la valeur d'ordre $\frac{N+1}{2}$
* Si N est pair, la médiane est la moyenne entre la valeur d'ordre $\frac{N}{2}$ et la valeur d'ordre $\frac{N+2}{2}$

Exemple 1:

- Soit un échantillon de 9 personnes dont le poids est :
  45 – 68 – 89 – 74 – 62 – 49 – 52 – 63 kg
  classés par ordre croissant :

$$\underbrace{45 - 49 - 52 - 56}_{4} - \underset{\text{médiane}}{\underset{\uparrow}{62}} - \underbrace{63 - 68 - 74 - 89}_{4} \quad \text{kg}$$

- Soit un échantillon de 10 personnes dont le poids est :
  45 – 68 – 89 – 74 – 55 – 62 – 49 – 52 – 63 kg
  classés par ordre croissant :

$$\underbrace{45 - 49 - 52 - 55 - 56}_{5} - \underbrace{62 - 63 - 68 - 74 - 89}_{5}$$

$$\text{médiane} = \frac{56+62}{2} = 59 \quad \text{kg}$$

Exemple 2:
Nombre d'enfants par famille

X_i	Effectifs f_i	Effectifs Cumulés F_i
0	3	3
1	8	11
2	12	23
3	5	28
4	2	30
Total	**N=30**	

N est pair ($\frac{N}{2}$ =15, $\frac{N+2}{2}$ =16)

0 0 0 1 1 1 1 1 1 1
1 2 2 2 **2** **2** 2 2 2 2
2 2 2 3 3 3 3 3 4 4

La médiane est $\frac{2+2}{2}$ =2

Pour déterminer la médiane d'un échantillon ou d'une population (variable continue) :

Si la variable est continue (regroupement par intervalle d'amplitude constante des résultats) le calcul de la médiane se fait autrement.
On utilise la colonne des effectifs cumulés pour déterminer la médiane
La médiane est le centre de l'intervalle correspondant a la première fréquence cumulée supérieure et égale à 0.5 , c'est-à-dire, 50 % de l'effectif total.

Exemple:
Résultats histoire-géographie dans une classe de 3ème

Classes	**Centre de la classe**	**Effectifs f_i**	**Effectifs cumulés F_i**	**Fréquences h_i**	**Fréquences cumulées H_i**
[0,2)	1	1	1	1/28	1/28
[2,4)	3	3	4	3/28	4/28
[4,6)	5	8	12	8/28	12/28=0,42
[6,8)	7	12	24	12/28	24/28=0,85
[8,10)	9	4	28	4/28	28/28=1
Total		**N=28**		**1**	

50% de N = 50% de 28=14

La médiane est 7

7 est le centre de l'intervalle [6,8) correspondant a la première fréquence cumulée supérieure et égale à 0,5. (Aussi, le centre de l'intervalle correspondant au premier effectif cumulé supérieur et égal à 14)

5. Les paramètres de position

Les **paramètres de position** sont, comme son nom l'indique, des valeurs de la variable (quantitative) qui régissent la position d'une donnée dans l'ensemble ordonné de valeurs.

DES QUANTILES

Les **quantiles** sont des valeurs qui divisent une série statistique ordonnée en plusieurs groupes comportant la même proportion de données.
Ce sont des valeurs de position, qui se calculent sur les mêmes principes que la médiane.

Les quantiles les plus couramment utilisés sont :

- Les **quartiles** (4 groupes) Q_1 ,Q_2 ,Q_3 Me=Q_2
- Les **déciles** (10 groupes) P_1, P_2…., P_{10} Me=P_5
- Les **centiles** (100 groupes) C_1, C_2…C_{100} Me=C_{50}

DES QUARTILES

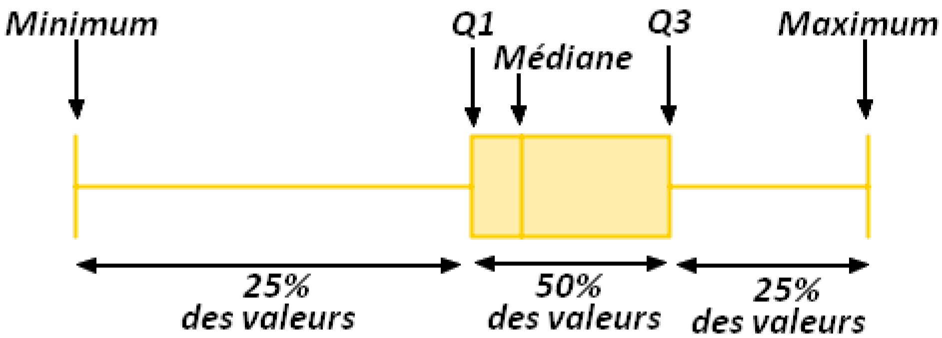
Les **quartiles** divisent une série statistique ordonnée, non plus en 2 comme la médiane, mais en 4 groupes comprenant chacun environ 25% des données de la série. Il y en a donc 3 : Q_1, Q_2 et Q_3
Alors :

- Avant le 1er quartile Q_1, on trouvera 25% des observations et, après lui, 75%.
- Avant le 2ème quartile Q_2, on trouvera 50% des observations et, après lui,
- 50%. Il s'agit de la médiane.
- Avant le 3ème quartile Q_3, on trouvera 75% des observations et, après lui, 25%.

L'intervalle interquartile $Q_3 – Q_1$ comprend 50% des observations en ayant éliminé les valeurs extrêmes.

Calcul des quartiles

- Calcul de N/4 pour Q_1 et 3N/4 pour Q_3
- Calcul des effectifs cumulés
- Lecture des quartiles à l'endroit où on atteint N/4 et 3N/4



Exemple:

Nombre d'enfants par ménage

X_i	Effectifs f_i	Effectifs cumulés F_i
0	150	150
1	250	**400**
2	300	**700**
3	150	**850**
4	100	950
5 et plus	50	1000
Total	**N=1000**	

Q_1= 1 enfant car N/4 = 1000/4 = **250**
Q_2= Mé= 2 enfants car N/2 = 2000/4 = **500**
Q_3= 3 enfants car 3N/4 = 3000/4 =**750**

Intervalle interquartile $Q_3 – Q_1$= 750-250=500

DES DÉCILES

les **déciles** correspondent aux valeurs de la variable qui partagent l'effectif total en 10 groupes d'effectifs identiques.
On calcule 9 déciles de D_1 à D_9 pour séparer la population en 10 parties égales.

Intervalle interdécile : $D_9 – D_1$

DES CENTILES

les **centiles** correspondent aux valeurs de la variable qui partagent l'effectif total en 100 groupes d'effectifs identiques.
On calcule 99 déciles de C_1 à C_{99} pour séparer la population en 100 parties égales.

6. Les paramètres de dispersion

L'étendue est la mesure de dispersion la plus simple à déterminer.
Elle mesure l'écart entre la plus grande et la plus petite valeur observée d'une série statistique quantitative.
E = Vmax - Vmin
Un instrument très grossier : L'étendue prend en compte les valeurs extrêmes d'une série statistique. Elle est donc très influencée par la présence d'une valeur aberrante très faible ou au contraire très élevée.

Exemple:
Temps pour accomplir une tâche (en secondes)

Série statistique	Étendue
4,4,5,7,8,8,9,10,11,6,7,12,5,8,5	12 – 4 = 8
4,4,5,7,8,8,9,10,11,6,7,12,5,8,5,25	25 – 4 = 21

L'individu le plus lent a mis 8 secondes de plus que l'individu le plus rapide pour accomplir la tâche demande dans la première série statistique, tandis que dans la seconde celui-ci a mis 21 secondes de plus.

Attention :
L'étendue et les quantiles ne prennent pas en compte l'ensemble des données pour mesurer la dispersion.

Une idée pour mesurer la dispersion:
On veut savoir si les N valeurs sont plus ou moins écartées de la moyenne. A chaque valeur x_i correspond un écart et on voudrait chiffrer l'ensemble de ces écarts.
Plus les écarts à la moyenne sont grands, plus la dispersion est forte.
On ne peut pas prendre la moyenne des différences (évidemment nulle)

- **Prendre la valeur absolue = Ecart moyen arithmétique**
- **Prendre le carré = Variance**

L'écart moyen arithmétique

Il s'agit d'une moyenne des écarts par rapport à la moyenne

- Calculer pour chaque modalité x_i son écart à la moyenne $|x_i - \bar{x}|$
- Faire la somme (pondérée) des écarts
- Diviser par l'effectif total

$$E_a = \frac{\sum_{i=1}^{N} f_{i\cdot}|x_i - \bar{x}|}{N}$$

Variance et écart-type

L'indice habituellement retenu est la moyenne des carrés des écarts à la moyenne, appelée la **variance**.

- Calculer pour chaque modalité x_i le carré de son écart à la moyenne $(x_i - \bar{x})^2$
- Faire la somme (pondérée) des écarts
- Diviser par l'effectif total

$$\sigma^2 = \frac{\sum_{i=1}^{N} f_{i\cdot}(x_i - \bar{x})^2}{N} = \frac{\sum_{i=1}^{N} f_{i\cdot} \cdot x_i^{\,2}}{N} - \bar{x}^2$$

On utilise la racine carrée de la variance, appelé **écart-type**, qui est exprimé dans la même unité de la variable x : $\sigma = \sqrt{\sigma^2}$

Le coefficient de variation

Lorsqu'on veut comparer deux séries par rapport à leur dispersion, alors que les observations ne sont pas exprimées dans la même unité (par exemple des euros et des dollars) ou lorsque les mesures ne sont pas dans les mêmes proportions (des dollars et des milliers de dollars), on utilise le rapport suivant, appelé « coefficient de variation » :

$$C_V = \frac{\sigma}{\overline{x}}$$

Il rapporte l'écart-type à la moyenne.
Plus grand est le coefficient de variation, plus grande est la dispersion.
On l'exprime souvent en **pourcentage**.

Remarque: Plus grand sont les paramètres de dispersion, plus grande est la dispersion. Par contre, si les paramètres sont petits, les donnés sont plus proche de la moyenne.

Exemple:
Résultats histoire-géographie dans une classe de 3ème

Classes	Centre classe X_i	Effectifs f_i	Effectifs cumulés F_i	$f_i \cdot X_i$	$f_i \cdot X_i^2$	$\lvert x_i - \overline{x}\rvert$	$f_i \cdot \lvert x_i - \overline{x}\rvert$
[0,2)	1	1	1	1	1	71/14	71/14
[2,4)	3	3	4	9	27	43/14	129/14
[4,6)	5	8	12	40	200	15/14	120/14
[6,8)	7	12	24	84	588	13/14	156/14
[8,10)	9	4	28	36	324	41/14	164/14
Total		**N=28**		**170**	**1140**		**640/14**

Étendue =10 – 0 = 10

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i \cdot f_i}{N} = \frac{1\cdot1+3\cdot3+5\cdot8+7\cdot12+9\cdot4}{28} = \frac{170}{28} = \frac{85}{14}$$

$$E_a = \frac{\sum_{i=1}^{5} f_{i\cdot}\lvert x_i - \overline{x}\rvert}{N} = \frac{640/14}{28} = \frac{80}{49} = 1{,}63$$

$$\sigma^2 = \frac{\sum_{i=1}^{5} f_{i\cdot} \cdot x_i^2}{N} - \overline{x}^2 = \frac{1140}{28} - \left(\frac{85}{14}\right)^2 = \frac{755}{196} = 3{,}85$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{755}{196}} = 1{,}96$$

$C_V = \frac{\sigma}{\overline{x}} = \frac{\sqrt{755/196}}{85/14} = 0{,}32326$ c'est-à-dire: 32,326%

7. L’analyse des paramètres statistiques

Les paramètres de tendance centrale et les paramètres de dispersion fournissent plus d’information quand on les étudie ensemble.

Les paramètres statistiques ont pour but de résumer, à partir de quelques nombres clés, l'essentiel de l'information relative à l'observation d'une variable quantitative.

Exemples: Regardez les pages 247, 248 et 249 du livre de l’élève.

Exercices du livre de l’élève. Pages 240-255