STATISTIQUES.

I. Médiane, quartiles et diagramme en boîte.

La médiane d'une série statistique partage la série en deux sous-séries de même effectif.

Soit une série statistique x_1 ; ; x_n dont les valeurs sont rangées dans **l'ordre croissant**.
Si la série comporte un nombre impair de valeurs (n est impair), la médiane est la valeur de rang $\frac{n+1}{2}$

Si la série comporte un nombre pair de valeurs (n est pair), la médiane est la moyenne des valeurs de rang $\frac{n}{2}$

et de rang $\frac{n}{2} + 1$.

Les quartiles partagent la série en quatre sous-séries :

Soit une série statistique x_1 ; ; x_n dont les valeurs sont rangées dans **l'ordre croissant**.
Le **premier quartile Q1** est la plus petite valeur x_i telles qu'au moins 25% des valeurs de la série soient inférieures ou égales à x_i.
Le **troisième quartile Q3** est la plus petite valeur x_i telles qu'au moins 75% des valeurs de la série soient inférieures ou égales à x_i.
L'écart interquartile est Q3 – Q1.
L'intervalle interquartile est l'intervalle [Q1 ;Q3].

On définit de même les déciles : le premier décile est le plus petit élément D1 de la série tel qu'au moins 10% des données soient inférieures à D1.
Le neuvième décile est le plus petit élément D9 de la série tel qu'au moins 90% des données soient inférieures à D9.

Exemples : Déterminer la médiane et les quartiles des séries ci-dessous :

Série n°1 : 2 ; 5 ;6 ;7 ;8 ;9 ;11 ;13 ;14 ;15 ;17 ;18 ;19 ;22 ;25 ;28
Il y a 16 valeurs.

La médiane est entre la 8ème et la 9ème valeur : Me = 13,5

$\frac{16}{4} = 4$ *donc Q1 est la 4ème valeur (dans l'ordre croissant) : Q1 = 7.*

$16 \times \frac{3}{4} = 12$ *donc Q3 est la 12ème valeur : Q3 = 18.*

Série n°2 : 1 ;5 ;6 ;7 ;8 ;9 ;15 ;16 ;17 ;18 ;21 ;25 ;29.
Il y a 13 valeurs.

La médiane est la 7ème valeur : Me = 15.

$\frac{13}{4} = 4{,}25$ *donc Q1 est la 5ème valeur : Q1 = 8.*

$13 \times \frac{3}{4} = 9{,}75$ *donc Q3 est la 10ème valeur : Q3 = 18.*

Série n°3 :

Le nombre de SMS reçus par des élèves d'un lycée au cours d'un week+ -end se répartit ainsi :

Nombre de SMS	0	5	12	15	17	18	25	50
Nombre d'élèves	18	10	15	12	8	5	10	1
effectifs cumulés	18	28	43	55	63	68	78	79

Il y a 79 valeurs.

La médiane est la 40ème valeur : Me = 12.

79/4 = 19,75 donc Q1 est la 20ème valeur : Q1 = 5.

$79 \times \frac{3}{4}$ = 59,25 donc Q3 est la 60ème valeur : Q3 = 17.

4, 5, 11, 13 page 66…

Remarques :

La médiane et les quartiles ne dépendent pas des valeurs extrêmes.
L'écart interquartile est un indicateur de dispersion de la série autour de la médiane. Il permet d'apprécier la dispersion des 50% des valeurs qui entourent la médiane. Plus il est petit, plus la série est homogène.

Diagramme en boîte.

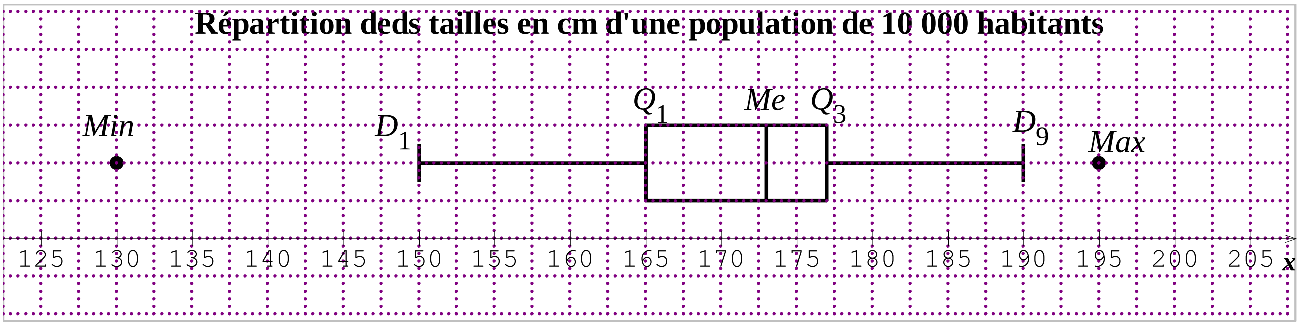
On construit un diagramme en boîte de la façon suivante :

☞ Les valeurs du caractère sont repérées sur un axe (vertical ou horizontal).

☞ On place sur cet axe le minimum et le maximum de la série, les quartiles, le 1er et le 9ème décile et la médiane.

☞ On construit un rectangle (boîte) parallèlement à l'axe dont les extrémités sont déterminés par les quartiles.

Exemple 1 : On donne le diagramme en boîte ci-dessous :



Lecture du diagramme en boîte :

- Les tailles minimales et maximales observées dans la population sont respectivement *130cm et 195cm*.
- D1 = 150 donc *10% de la population mesure moins de 150 cm*
- D9 = 190 donc *10 % mesure plus de 190cm.*
- Q1 = 165 donc *25% de la population mesure moins de 165 cm*
- Q3 = 177 donc *25 % mesure plus de 177cm.*
- Me = 173 donc *50% mesure moins de 173 cm.*

Exemple 2 :
Construire le diagramme en boîte de la série 3 ci-dessus.

6, 14, 15, 16 page 66…

II. Moyenne, variance et écart type.

Dans tout le paragraphe, on considère la série statistique définie par le tableau ci-dessous :

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p

$N = n_1 + n_2 + \ldots + n_p$ est l'effectif total

1. Un paramètre de tendance centrale : la moyenne.

Définition : La moyenne de la série statistique est le nombre réel, noté $\overline{x}$ défini par

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$$

***Remarque* : Lorsque la série est regroupée en classes, on prend pour valeurs des x_i les centres des classes pour obtenir une valeur approchée de la moyenne.**

Exemple : On reprend la série 3 plus haut :

Nombre de SMS	0	5	12	15	17	18	25	50
Nombre d'élèves	18	10	15	12	8	5	10	1

Le nombre moyen de SMS envoyés par élève est environ 11,85.

2. Des paramètres de dispersion : variance et écart type.

Définition : L'écart type de la série est le nombre positif, noté σ, défini par

$$\sigma = \sqrt{\frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \ldots + n_p(x_p - \overline{x})^2}{N}}.$$

σ est dans la même unité que les valeurs de la série.
L'écart type permet de mesurer la dispersion de la série. Plus il est grand, plus la série est dispersée.
On l'obtient à la calculatrice, où il est noté σx ou $x\sigma n$

Exemple : Dans l'exemple 3 ci-dessus, on obtient $\sigma \approx$ *9,2*

20, 23, 24, 25, 32 page 66…

III. Résumé d'une série statistique.

Si l'on souhaite retenir un seul nombre pour résumer une série statistiques, on choisit un indicateur de position : la médiane ou la moyenne.
Si l'on souhaite aussi rendre compte de la répartition des valeurs autour de cette "valeur centrale", on lui associe un indicateur de dispersion : l'écart interquartile ou l'écart type.
On obtient ainis deux couples (**moyenne ; écart type**) et (**médiane ; écart interquartile**) qui sont deux résumés d'une série statistique. Ils ne prétendent pas restituer toute l'information de la série statistique mais ils permettent d'en synthétiser l'essentiel et de faciliter la comparaison entre plusieurs séries.

1. Le couple (médiane ; écart interquartile)

Ce coupe donne à la fois :

- un indicateur de tendance centrale : la médiane
- un indicateur de dispersion : la longueur de l'intervalle interquartile qui contient la moitié des valeurs de la série.

Plus l'écart interquartile est petit, plus les valeurs centrales de la série se concentrent autour de la médiane.

2. Le couple (moyenne ; écart type)

Ce coupe donne à la fois :

- un indicateur de tendance centrale : la moyenne
- un indicateur de dispersion : l'écart type

Plus l'écart type est petit, plus les valeurs se concentrent autour de la moyenne.

3. Choix du couple.

Le couple (médiane ; écart interquartile) :

il est assez facile à interpréter

il est très peu sensible aux valeurs extrêmes (parfois suspectes)

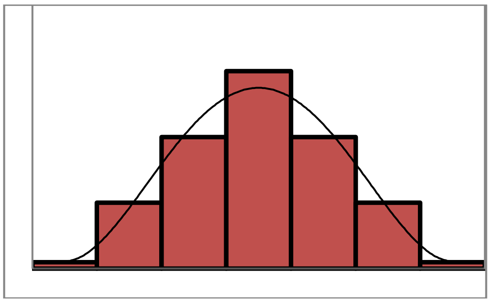
Mais il ne se calcule pas par paquets : connaissant la médiane de de sous-séries constituant une série, on ne peut pas calculer la médiane de cette série.

Le couple (moyenne ; écart type) :

il joue un grand rôle en statistique théorique (par exemple appliquée aux sondages)

les définitions algébriques de la moyenne et de l'écart type permettent d'obtenir différentes propriétés

Mais l'écart type tient compte des écarts de toutes les valeurs à la moyenne. Il donne beaucoup de poids aux valeurs extrêmes et son choix n'est pertinent que lorsque le diagramme en barre qui représente la série est assez symétrique et évoque la forme d'une courbe en cloche comme ci-dessous :



IV. Effet de structure.

En 2015, le personnel d'une petite entreprise était constitué par 3 cadres et par 7 ouvriers. Le salaire mensuel d'un cadre était alors de 10 000€ et celui d'un ouvrier de 1 000€.

En 2016, l'entreprise compte 2 cadres et 8 ouvriers et le salaire mensuel d'un cadre est de 10 100€ et celui d'un ouvrier de 1 100€.

Le PDG de l'entreprise se félicite : "les cadres et les ouvriers ont été augmentés cette année !".

Anatole, un des ouvriers, affirme que le salaire moyen dans cette entreprise a diminué. Qu'en pensez-vous?

Le PDG a raison car 10 100 > 10 000 et 1 100 > 1 000

Salaire moyen en 2015 : 3 700€

Salaire moyen en 2016 : 2 900€ donc Anatole a raison aussi.

Malgré une hausse de tous les salaires, le salaire moyen a baissé car la structure de l'entreprise a changé. C'est l'effet de structure.