Première S	**Statistiques**	Année scolaire 2012/2013

I) Rappels de seconde :

1) Médiane d'une série statistiques :

On considère une série statistique (x_i) dont les valeurs sont rangées dans l'ordre croissant.
On suppose que cette série statistique comporte N valeurs.
La médiane partage cette population en deux ensembles de même effectif.
(= il y a autant de valeurs inférieures à la médiane que de valeurs supérieures)

- Si N est impair, alors la médiane est la $\frac{N+1}{2}$ ième valeur
- Si N est pair, alors la médiane est la moyenne de la $\frac{N}{2}$ et de la $\frac{N}{2}+1$ ième valeur.

2) Premier quartile :

C'est la plus petite valeur de la série telle qu'au moins 25% des valeurs lui sont inférieures ou égales.

Méthode :
Après avoir rangé les valeurs prises par le caractère dans l'ordre croissant (en les écrivant autant de fois qu'elles apparaissent), on calcule le quart de l'effectif.
Soit on tombe sur un entier alors il désigne la position du premier quartile, soit on prend le terme qui suit.

Exemple :
Voici les notes obtenues à un devoir de mathématiques par une classe de 30 élèves :

Notes	5	7	8	9	10	12	13	14	15	18	19
Effectifs	2	3	2	6	4	5	1	3	2	1	1

Premier quartile : il y a 30 valeurs en tout.
Or, $30 \times \frac{1}{4} = 7{,}5$. On prend donc la 8 ième note , c'est-à-dire : 9

Donc le premier quartile est égal à 9

3) Troisième quartile :

C'est la plus petite valeur de la série telle qu'au moins 75% des valeurs lui sont inférieures ou égales.

Dans l'exemple précédent, $30 \times \frac{3}{4} = 22{,}5$. On prend alors la 23ième note, c'est-à-dire : 13. Donc le troisième quartile est égal à 13

4) Etendue :

C'est l'écart entre la plus petite valeur du caractère et la plus grande.
Dans l'exemple précédent : 19 – 5 = **14**

5) Intervalle interquartile :

Si on note Q_1 et Q_3, respectivement premier et troisième quartile, alors l'intervalle interquartile est: **[Q1 ; Q3]**

6) Ecart interquartile :

C'est $Q_3 - Q_1$.

Il constitue une mesure de la dispersion liée à la médiane.

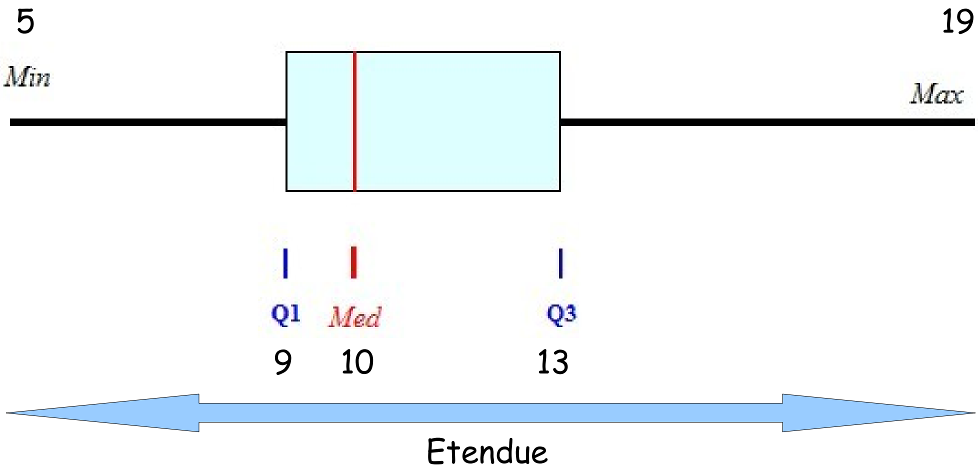
II) Diagramme en boîte :

(Box plot ou diagramme de Tukey (statisticien américain (1915-2000))

On peut représenter sur un même graphique les deux quartiles, la médiane, l'étendue, le minimum et le maximum.

On appelle ces représentations diagrammes en boîtes, ou boîtes à moustaches ou encore diagramme de Tukey)

Celui qui correspond à l'exemple précédent est :



III) Variance et écart-type :

1) Variance :

Soit p un entier naturel non-nul.

On considère une série statistique $x_1, x_2, x_3, \dots, x_p$ et les effectifs correspondants : $n_1, n_2, n_3, \dots, n_p$

On appelle fonction de dispersion la fonction d définie pour tout x par :

$$d(x) = n_1(x_1 - x)^2 + n_2(x_2 - x)^2 + \dots + n_p(x_p - x)^2$$

En développant cette expression :

$d(x) = n_1(x_1^2 - 2x_1x + x^2) + n_2(x_2^2 - 2x_2x + x^2) + \dots + n_p(x_p^2 - 2x_px + x^2)$

Ensuite, on rassemble les termes en x^2, ceux en x puis ceux ne contenant pas de x :

$d(x) = (n_1 + n_2 + \dots + n_p)x^2 - 2x(n_1x_1 + n_2x_2 + \dots + n_px_p) + n_1x_1^2 + n_2x_2^2 + \dots + n_px_p^2$

d est un trinôme du second degré dont l'expression en fonction de x est $ax2 + bx + c$ avec $a = n_1 + n_2 + \dots + n_p > 0$ car les effectifs sont tous positifs.

D'où d est d'abord décroissante , puis ensuite croissante.

Pour trouver l'abscisse de son minimum, on calcule d' :

$$d'(x) = 2(n_1 + n_2 + \dots + n_p)x - 2(n_1x_1 + n_2x_2 + \dots + n_px_p)$$

Or, $d'(x) = 0 \Leftrightarrow 2(n_1 + n_2 + \dots + n_p)x - 2(n_1x_1 + n_2x_2 + \dots + n_px_p) = 0$

$\Leftrightarrow x = \dfrac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p}$ = Moyenne des x_i

Autrement dit, la fonction de dispersion atteint son minimum pour x = moyenne des x_i

Le minimum atteint s'appelle **la variance de la série statistique.**

Définition :

Soit p un entier naturel non-nul :

On considère une série statistique x_1, x_2, x_3, … , x_p et les effectifs correspondants : n_1, n_2, n_3, … , n_p . Notons N = effectif total = $n_1 + n_2 + \ldots + n_p$

Alors, la variance de cette série statistique notée V est définie par :

$$\boxed{V = \frac{1}{N}\left(n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \ldots + n_p(x_p - \overline{x})^2\right)}$$

(si on note $\overline{x}$ la moyenne des x_i)

2) Ecart-type :

L'écart-type est la racine carrée de la variance.

Il se note s en statistiques et s = $\sqrt{V}$

Il mesure la dispersion liée à la moyenne.

3) Utilisation de la calculatrice :