Dans la suite, x_i, n_i et $f_i$désignent respectivement le ième caractère, son effectif et sa fréquence.

I Médiane, quartiles et diagramme en boîte

1.1 Médiane et quartile

Définition La médiane est la valeur (notée Me) qui partage la série en deux populations d'effectif égal. Pour la déterminer, si l'effectif N est impair (respectivement pair) alors la médiane est la valeur de la série de rang $\frac{N+1}{2}$ (respectivement est la moyenne des valeurs de rang $\frac{N}{2}$ et $\frac{N+2}{2}$).

Attention ! On range la série dans l'ordre croissant.

Définition Le 1er quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs sont inférieurs ou égales à Q_1. Le 3ème quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs sont inférieurs ou égales à Q_3.

Attention ! On range la série dans l'ordre croissant.

Remarques 1) La médiane et les quartiles sont des paramètres de tendance centrale et de position.
2) La médiane et les quartiles sont des paramètres qui ne sont pas sensibles aux valeurs extrêmes.

Exemple Voici une liste de températures (en C°) relevées sous abri à différents moments d'une journée ; les données sont rangées dans l'ordre croissant :
3 - 3.8 - 4.5 – 4.8 - 5 – 5.5 – 5.7 – 5.8 - 6.2 – 7 – 7.3 – 8.2 – 9 – 9.2 – 9.5 – 9.7
Ici N=16, donc la médiane est : $\frac{5{,}8+6{,}2}{2} = \frac{12}{2} = 6$. Pour déterminer Q_1, on fait $\frac{1}{4}N = 4$; donc $Q_1 = 4{,}8$. Pour déterminer Q_3, on fait $\frac{3}{4}N = 12$; donc $Q_3 = 8{,}2$.

1.2 L'écart interquartile

Définition L'écart interquartile est la différence $E_I = Q_3 - Q_1$ entre le 3ème quartile et le 1er quartile de la série.

Ce paramètre mesure la dispersion autour de la médiane (il ignore les valeurs extrêmes de la série statistique).

Exemple Reprenons l'exemple précédent : $E_I = Q_3 - Q_1 = 8{,}2 - 4{,}8 = 3{,}4$.

1.3 Diagramme en boîte

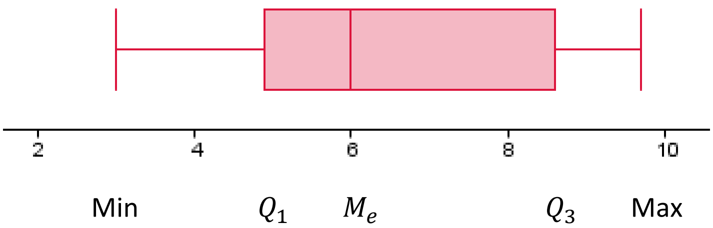
Il résume graphiquement une série statistique et l'on retrouve les paramètres précédents. On l'appelle aussi boîte à moustache ou diagramme de Tukey.

Histoire des mathématiques *John Wilder Tukey*
New. Bedford (Massachusetts) 1915- Princeton (New Jersey) 2000
Les parents du mathématicien américain John TUKEY, enseignants dans le secondaire, n'envoient pas leur fils à l'école et se chargent de son instruction. Il entame des études de chimie, et entre à l'université de Princeton en 1936. Il y étudie parallèlement les mathématiques, et c'est dans cette discipline qu'en 1939, il obtient son doctorat, sous la direction de LEFCHETZ, pour une thèse sur *la dénombrabilité en topologie*. Durant la Seconde Guerre

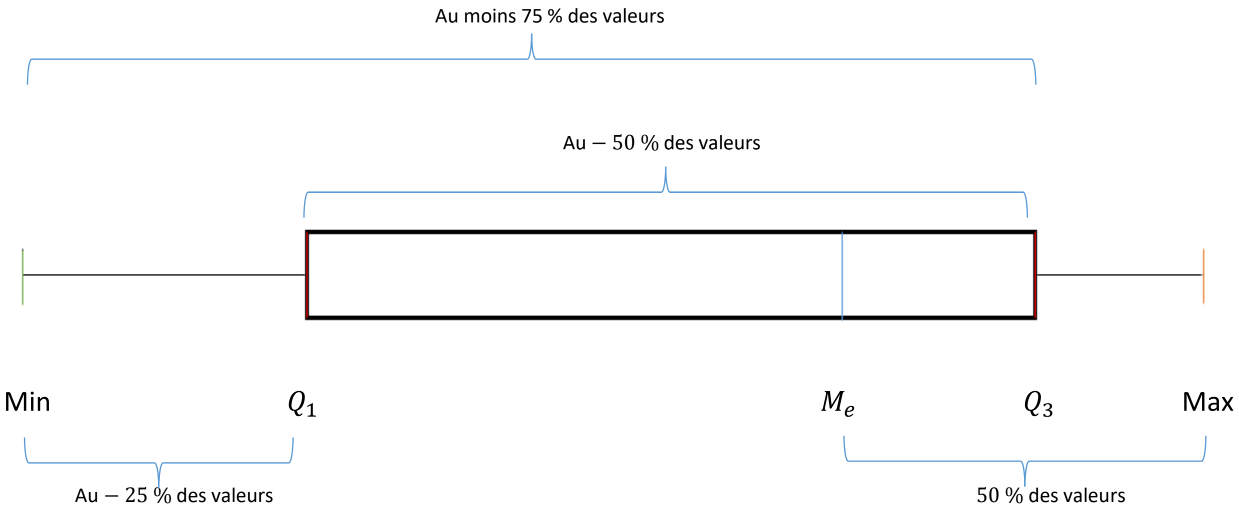
mondiale, TUKEY est employé dans un bureau stratégique de recherches mathématiques, où il aborde les statistiques : sa vocation est née ! En 1945, il obtient un poste dans cette discipline à Princeton, université où il effectuera toute sa carrière. Parmi les nombreuses méthodes statistiques nouvelles introduites par TUKEY, on peut citer *l'algorithme de la transformée rapide*, qu'il développe en 1965 avec James COOLEY, mathématicien travaillant pour la firme IBM. Il recentre la recherche en statistiques sur sa mission initiale d'analyse mathématique des données, ce qui le conduit à travailler sur différents problèmes concrets, comme le recensement de la population, les sondages électoraux, la destruction de la couche d'ozone par les aérosols. Sa passion de l'informatique, à la fin de la guerre, l'amène à créer le mot valise *bit,* pour *binary unit*, terme qui a fait depuis le tour du monde, ainsi que le mot *software*, qui désigne les logiciels en anglais.

Citation : « une réponse approximative à une bonne question est bien préférable à une réponse précise à une mauvaise question .»

D'après l'exemple précédent, on obtient le diagramme de Tukey suivant :



BILAN



Remarque Ce diagramme est surtout utilisé pour comparer la répartition des données de plusieurs séries (Températures d'une ville sur deux périodes, tailles de deux adolescents, prix de deux actions…).

II Variance et écart type

Dans ce paragraphe, on va mettre en place d'autres paramètres qui vont nous permettre de dire si les valeurs de la série sont regroupées ou au contraire dispersées (autour de sa moyenne et entre elles).
On considère dans la suite, la série statistique définie par le tableau ci-dessous.

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p

L'effectif total est $N = n_1 + n_2 + \cdots + n_p$.

2.1 Un rappel : formule de la moyenne

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

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_p n_p}{N} = \frac{1}{N} \sum_{i=1}^{p} n_i x_i$$

Remarque La moyenne est un paramètre de tendance centrale.

2.2 Définitions

Définition La variance de la série statistique est le nombre réel, noté V, défini par :

$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \cdots + n_p \times (x_p - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^{p} n_i (x_i - \bar{x})^2 .$$

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

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

Interprétation La variance est la moyenne des carrés des écarts à la moyenne. Elle mesure donc la dispersion autour de la moyenne. Elle n'est pas très parlante car elle s'exprime dans le carré de l'unité du caractère.
L'écart-type a l'avantage de s'exprimer dans la même unité que le caractère. Il permet de comparer la dispersion de la série de valeurs (par rapport à la moyenne) et l'intérêt réside dans la comparaison de valeurs. Contrairement à l'écart interquartile, il tient compte de l'ensemble de la population (les valeurs extrêmes influencent l'écart-type).

Exemples
Voici les notes de mathématiques d'Anna et Brahim obtenues lors des 4 derniers DS :
Anna : 8 – 9 – 11 – 9
Brahim : 1 - 14 – 3 – 19
Comparons ces deux séries de notes :

- Anna : $\overline{x_1} = 9{,}25$; $V_1 = 1{,}19$, $\sigma_1 = 1{,}09$

- Brahim : $\overline{x_2} = 9{,}25$; $V_2 = 56{,}19$, $\sigma_2 = 7{,}5$

$\overline{x_1} = \overline{x_2}$ mais $\sigma_1 < \sigma_2$. Les notes de Brahim sont plus dispersées par rapport à leur moyenne.