

Etude des séries statistiques

Classe de première ES - Lycée Saint-Charles

Patrice Jacquet - www.mathxy.fr - 2013

Objectifs :

- Connaître les indicateurs de tendance centrale.
- Connaître les indicateurs de dispersion.
- Savoir interpréter un diagramme en boîte.

1 Médiane, quartile et diagramme en boîte

1.1 Paramètres de tendance centrale : médiane et quartiles

Définition 1 – Médiane

Une **médiane** d'un ensemble de valeurs est une valeur qui permet de couper l'ensemble en deux parties égales.

Méthode - Pour déterminer une médiane d'un ensemble de valeurs, il suffit d'ordonner les valeurs en une liste croissante et de choisir la valeur qui est au centre de cette liste.

Exemple 1 – Soit un ensemble de 7 entiers : $\{10, 7, 9, 73, 1, 2013, 0\}$.

Après tri, la série est 0, 1, 7, 9, 10, 73, 2013. La médiane est le quatrième élément de cette série : $M_e = 9$. Quatre valeurs de l'ensemble sont inférieures ou égales à 9, et quatre sont supérieures ou égales à 9.

Exemple 2 – Soit un ensemble de 6 entiers : $\{10, 7, 9, 1, 9, 0\}$.

Après tri, la série est 0, 1, 7, 9, 9, 10. Toute valeur comprise entre 7 et 9 peut être choisie comme une médiane. On prend généralement la valeur moyenne, c'est à dire $M_e = 8$ dans cet exemple.

Définition 2 – Quartiles

Les **quartiles** d'un ensemble de valeurs sont chacune des trois valeurs qui divisent les données triées en quatre parties égales.

Le **premier quartile** Q_1 sépare les 25% inférieurs des données (au moins un quart des données sont **inférieures ou égales à** Q_1).

Le deuxième quartile est la médiane.

Le **troisième quartile** Q_3 sépare les 75% inférieurs des données (au moins les trois quart des données sont **inférieures ou égales à** Q_3).

Exemple 3 – Soit un ensemble de 9 entiers : $\{10, 7, 9, 1, 9, 0, 12, 24, 15\}$.

Après tri, la série est 0, 1, 7, 9, 9, 10, 12, 15, 24.

$9 : 4 = 2,25$ donc Q_1 est la troisième valeur : $Q_1 = 7$.

$3 \times (9 : 4) = 6,75$ donc Q_3 est la septième valeur : $Q_3 = 12$.

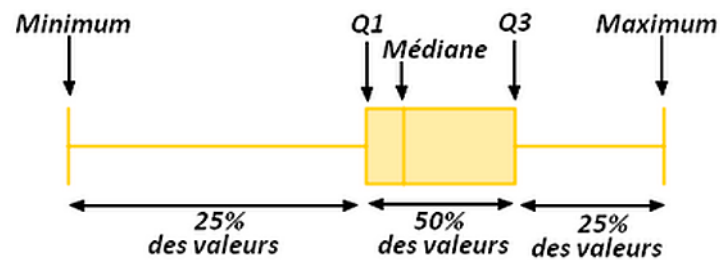
1.2 Paramètres de dispersion : écart interquartile

Définition 3 – Écart interquartile

L'écart interquartile est la différence entre le troisième quartile et le premier quartile : $Q_3 - Q_1$.

1.3 Diagramme en boîte (ou boîte à moustache)

Le diagramme en boîte (aussi appelée boîte à moustaches) est un moyen rapide de visualiser la répartition des données autour de la médiane. Ce diagramme est souvent utilisé pour comparer un même caractère dans deux populations de tailles différentes (voir activité 2).



2 Moyenne, variance et écart-type

Dans toute la suite on considèrera la série statistique :

Valeur	x_1	x_2	...	x_p	TOTAL
Effectif	n_1	n_2	...	n_p	$N = n_1 + n_2 + \dots + n_p$

2.1 Indicateur de tendance centrale : moyenne

Définition 4 – Moyenne

La moyenne \bar{x} de cette série est le nombre défini par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}.$$

On peut aussi écrire :

$$\bar{x} = \sum_{i=1}^p \frac{n_ix_i}{N} \quad (\text{somme pour } i = 1 \text{ à } i = p \text{ des nombres } \frac{n_ix_i}{N}).$$

2.2 Paramètres de dispersion : variance et écart-type

Définition 5 – Variance

La variance V de cette série est le nombre défini par :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}.$$

On peut aussi écrire :

$$V = \sum_{i=1}^p \frac{n_i(x_i - \bar{x})^2}{N}.$$

Définition 6 – Écart-type

L'écart-type σ de cette série est le nombre défini par : $\sigma = \sqrt{V}$

3 Résumé d'une série statistique

On résume une série statistique par un paramètre de tendance centrale (**médiane** ou **moyenne**) associé à un paramètre de dispersion (**écart interquartile** ou **écart-type**).

Le **couple (moyenne ; écart-type)** a l'inconvénient d'associer deux paramètres sensibles aux valeurs extrêmes, contrairement au **couple (médiane ; espace interquartile)**.

Mathématiquement la notion d'écart-type est fondamentale. Les écarts-types sont rencontrés dans tous les domaines où sont appliqués les probabilités et la statistique, en particulier dans le domaine des **sondages**, en **physique**, en **biologie** ou dans la **finance**.