Olivier FINOT, 21 août 2015

I. Indicateurs de tendance centrale

1) Mode

Définition

Le **mode** d'une série statistique est la valeur qui a **l'effectif le plus important**.

Exemple

Le tableau suivant présente le nombre de repas pris chaque semaine par les élèves d'un lycée professionnel :

Nombre de repas	0	1	2	3	4	5
Nombre d'élèves	56	24	72	99	259	115

Ici le **mode** est 4, car $259 > 56$ et $259 > 24$ et $259 > 72$ et $259 > 99$ et $259 > 115$.

2) Médiane

Définition

La **médiane** Me d'une série statistique est le nombre qui **partage la série en deux** séries ayant **le même effectif**.
La moitié (ou 50 %) des valeurs de la série sont inférieures ou égales à la médiane et l'autre moitié (50 %) lui sont supérieures ou égales.

Méthode

Pour calculer la valeur Me de la médiane d'une série statistiques :

- ranger les valeurs par ordre croissant (du plus petit grand) ;
- → si l'effectif total (N) est impair, Me est la $\left(\frac{N+1}{2}\right)^e$ valeur de la série.

  → si N est pair, Me est la moyenne entre les $\left(\frac{N}{2}\right)^e$ et $\left(\frac{N}{2}+1\right)^e$ valeurs.

Exemple

Le 1^{er} novembre 2012, on a relevé le prix du gazole sur 10 points de vente du département du Territoire de Belfort. Les 10 prix rangés dans l'ordre croissant sont :

Rang	1	2	3	4	5	6	7	8	9	10
Prix	1,368	1,369	1,374	1,375	1,377	1,379	1,385	1,408	1,450	1,460

Ici $N = 10$ est pair.
La **médiane** Me de la série est donc la moyenne entre les 5^e et 6^e valeurs :

$$Me = \frac{n_5 + n_6}{2} = \frac{1,377 + 1,379}{2} = 1,378$$

La moitié des prix pratiqués est donc inférieure ou égale à 1,378 €.

3) Moyenne

Définition

On note $x_1, x_2, ..., x_p$ les valeurs du caractère étudié et $n_1, n_2, ..., n_p$ les effectifs correspondants.
La **moyenne** $\bar{x}$ de la série statistique est $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + ... + n_px_p}{N} = \dfrac{\Sigma n_i x_i}{N}$

Exemple

Ici on considère la répartition des prix du gazole dans l'ensemble des 25 stations du département :

Prix	1,368	1,369	1,374	1,375	1,377	1,379	1,385	1,408	1,450	1,460
Nb. de stations	2	5	2	4	1	4	2	1	3	1

Moyenne des prix des 25 stations :

$$\bar{x} = \frac{1,368 \times 2 + 1,369 \times 5 + ... + 1,450 \times 3 + 1,460}{25} = 1,3884$$

Le prix moyen observé pour ces 25 stations est 1,3884 €.

Ne pas confondre moyenne et médiane qui sont deux indicateurs statistiques très différents.

En 2013 le salaire moyen français était de 7.800 € bruts par mois, alors que le salaire médian était de 1.675 € bruts par mois [a].

[a] Source : http ://www.lefigaro.fr/social/2014/03/13/09010-20140313ARTFIG00167-derriere-le-salaire-moyen-de-fortes-disparites.php

II. Indicateurs de dispersion

1) Étendue

Définition

L'**étendue** e d'une série statistique est la différence entre la plus grande et la plus petite valeur de la série.

Exemple

Les 10 prix rangés par ordre croissant sont :

Rang	1	2	3	4	5	6	7	8	9	10
Prix	1,368	1,369	1,374	1,375	1,377	1,379	1,385	1,408	1,450	1,460

L'étendue de la série est $e = 1,460 - 1,368 = 0,092$ €.

2) Quartiles

Définition

- Le **premier quartile** Q_1, est la plus petite valeur à laquelle un quart (ou 25 %) des valeurs sont inférieures ou égales.
- Le **troisième quartile** Q_3, est la plus petite valeur à laquelle trois quarts (ou 75 %) des valeurs sont inférieures ou égales.
- L'**écart interquartile** $Q_3 - Q_1$ est la différence entre les 3^e et 1er quartiles : $Q_3 - Q_1$. Il regroupe au moins 50 % des effectifs de la série avec un nombre égal de valeurs réparties de part et d'autre de la médiane Me.

Méthode

Pour calculer les valeurs Q_1 et Q_3 des quartiles :

- ranger les valeurs de la série par ordre croissant ;
- calculer $r_1 = 0,25 \times N$ et $r_3 = 0,75 \times N$;
- → Si N est un multiple de 4, Q_1 est la r_1^e valeur de la série et Q_3 est la r_3^e valeur de la série.

  → Si N n'est pas un multiple de 4,

  — le plus petit entier supérieur à r_1 (ou $\lceil r_1 \rceil$) donne le rang de Q_1.

  — le plus petit entier supérieur à r_3 (ou $\lceil r_3 \rceil$) donne le rang de Q_3.

Exemple

- Dans l'exemple ci-dessus, on a $N = 10$, donc N n'est pas un multiple de 4.
  $r_1 = 0,25 \times N = 0,25 \times 10 = 2,5$ et $r_3 = 0,75 \times N = 0,75 \times 10 = 7,5$
- Calcul du `premier quartile` Q_1 :

  → le plus petit entier supérieur à $r_1 = 2,5$ est 3 ;

  → Q_1 correspond à la 3^e valeur de la série : $Q_1 = 1,374$
- Calcul du `troisième quartile` Q_3 :

  → le plus petit entier supérieur à $r_3 = 7,5$ est 8 ;

  → Q_3 correspond à la 8^e valeur de la série : $Q_3 = 1,408$

  25 % des prix pratiqués sont inférieurs ou égaux à 1,374 € et 75 % des prix pratiqués sont inférieurs à 1,408 €.
- L'`écart interquartile` $Q_3 - Q_1$ vaut $1,408 - 1,374 = 0,034$ €.

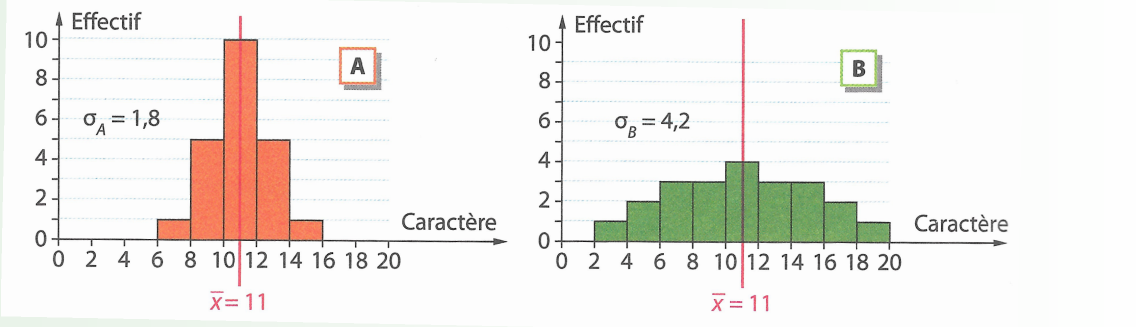
3) Écart type

Définition

L'`écart type` σ (sigma), fourni par la calculatrice ou le tableur, mesure la dispersion de la série autour de la moyenne $\bar{x}$.
Plus l'écart type σ est grand, plus les valeurs sont «`dispersées`» autour de la moyenne.
Inversement, plus l'écart type σ est grand, plus les valeurs sont «`resserrées`» autour de la moyenne.

Exemple

Ces deux graphiques représentent deux séries de même effectif et de de même moyenne $\bar{x} = 11$.



$\sigma_A < \sigma_B$: les valeurs de la série **B** sont plus dispersées que celles de la série **A** autour de $\bar{x}$.

III. Diagrammes en boîte à moustaches

Définition

Le **`diagramme en boîte à moustaches`** est un dessin à l'échelle, où la «**`boîte`**» est un rectangle limité par Q_1 et Q_3, et regroupe donc 50 % des valeurs.
La médiane Me est repérée par un segment dans le rectangle.
Le minimum x_{min} et le maximum x_{max} correspondent aux extrémités des «**`moustaches`**».

Exemple

Boite à moustache correspondant à l'exemple :