

Chapitre 5. Statistiques

Yann Barsamian

École Européenne de Bruxelles 1

Année scolaire 2022–2023



- Vocabulaire de base
- Caractéristiques de séries statistiques
- Représentations de séries statistiques

On va analyser des données brutes. Ces données peuvent être quantitatives (ce sont des nombres, par ex. votre meilleur temps au 100 m en course à pied) ou qualitatives (ce ne sont pas des nombres, par ex. votre matière préférée l'an dernier).

Pour avoir une vue d'ensemble de la série statistique (l'ensemble des données), on peut calculer les effectifs de chaque valeur : c'est le nombre de fois que la valeur apparaît dans la série. L'effectif total est le nombre total de valeurs.

Exemple : une étude sur 20 élèves concernant le temps de travail journalier (en minutes) à la maison donne les résultats suivants :

Temps	5	15	20	30	40	60
Effectif	1	4	3	6	5	1

Remarque : les valeurs 5 et 60 sont des valeurs aberrantes pour la série (très différentes des autres valeurs).

II/ Caractéristiques d'une série statistique

1) La moyenne¹ :

C'est la somme des valeurs divisée par le nombre de valeurs. Cela revient à répartir le total pour que chacun ait la même chose.

Exemple : la moyenne de 8, 9 et 13 est $\frac{8 + 9 + 13}{3} = 10$.

Avec des coefficients (ou des effectifs), il faut pondérer : on multiplie chaque valeur par l'effectif avant de diviser par l'effectif total. Voir <https://www.youtube.com/watch?v=Rhgv1gRUI2w>, ou l'exemple :

Temps	5	15	20	30	40	60
Effectif	1	4	3	6	5	1

$$\frac{1 \cdot 5 + 4 \cdot 15 + 3 \cdot 20 + 6 \cdot 30 + 5 \cdot 40 + 1 \cdot 60}{20} = \frac{565 \div 5}{20 \div 5} = \frac{113}{4} = 28,25.$$

Ainsi, les élèves travaillent en moyenne 28 minutes par jour.

1. https://www.youtube.com/watch?v=88_16UbkdZM

2) La médiane² :

C'est un nombre qui sépare la série en deux parties de même effectif. Au moins 50% des valeurs sont inférieures ou égales à la médiane, et au moins 50% des valeurs sont supérieures ou égales à la médiane.

Pour calculer la médiane d'une série de n nombres, on commence par ordonner les valeurs si ce n'est pas déjà fait :

$$u_1 \leq u_2 \leq \dots \leq u_n$$

- si n est impair, c'est la valeur centrale

Ex. : 2, 5, 7, 8, 9 : la médiane est 7

- si n est pair, il y a deux valeurs centrales ;

Ex. : 2, 5, 7, 8, 9, 12 : la médiane est $\frac{7+8}{2} = 7,5$ (en fait toute valeur entre 7 et 8 convient).

2. <https://www.youtube.com/watch?v=g1OCTw--VYQ>

3) Le mode (ou valeur dominante) :

C'est la valeur la plus présente dans la série.

Sur notre exemple précédent :

Temps	5	15	20	30	40	60
Effectif	1	4	3	6	5	1

Le mode est 30 (parfois il peut y avoir plusieurs modes si plusieurs valeurs ont l'effectif maximum).