

STATISTIQUES

MATHÉMATIQUES 4G

5 mai 2021

1 Série statistique

Une série statistique est une série d'observations, de mesures, de relevés (appelés données ou DATA)

Exemples :

- les âges des élèves de rhéto de la région bruxelloise (sciences sociales)
- les noms les plus utilisés par un internaute lors d'une recherche GOOGLE (marketing)
- les poids des camemberts PRESIDENT dans un lot (industrie)
- les températures quotidiennes du mois de mai (météorologie)
- les intentions de vote dans une ville aux prochaines élections (politique/sondage)
- les calories alimentaires dépensées au quotidien par les habitants de Bruxelles (santé)

L'objectif est d'analyser ces DATA afin d'en tirer des tendances, des conclusions. On note N le nombre total d'observations. Ces observations sont, le plus souvent, nombreuses et on range donc d'abord dans des tableaux.

2 Les tableaux statistiques

Par exemple, on étudie la couleur préférée (=le caractère observé) des élèves de quatrième (=la population). Le titre du graphique doit reprendre le caractère observé et la population considérée.

Les intitulés des colonnes sont :

- les valeurs des caractères observés triés dans l'ordre s'il s'agit de nombres (éventuellement regroupés en intervalles, appelés classes)
- les effectifs : le nombre de fois qu'on a observé chaque valeur du caractère
- la fréquence : le pourcentage avec lequel chaque valeur du caractère a été observée
- les fréquences cumulées : voir explication dans l'exercice

➤ des colonnes supplémentaires qui permettent d'effectuer certains calculs

Exercice collectif : créer un tableau de distribution qui correspond à une note d'autoévaluation (sur une échelle de 1 à 10) sur le niveau en mathématiques des élèves de la classe

3 Les indicateurs

Les indicateurs sont des outils mathématiques qui permettent d'analyser les DATA à partir du tableau de distribution. Dans cette section, nous décrivons quelques-uns d'entre-eux. Calculons ensuite ces indicateurs pour notre exemple précédent (exercice).

3.1 L'étendue E

L'étendue est la différence entre la plus grande valeur observée et la plus petite.

$$E = x_{max} - x_{min}$$

3.2 Le mode

Le mode est la (ou les) valeur(s) du caractère dont l'effectif est le plus grand.

3.3 La moyenne \bar{x}

La moyenne est la somme de toutes les valeurs observées, divisée par le nombre total d'observations.

$$\bar{x} = \frac{\sum_i n_i \cdot x_i}{N}$$

3.4 La médiane M

La médiane est une valeur telle qu'il y ait autant de valeurs inférieures que supérieures à cette valeur.

Pour déterminer la médiane :

➤ on range les valeurs par ordre croissant

➤ si N est impair : la médiane est la valeur se situant à la position $\frac{N+1}{2}$

3.5 Le premier quartile Q_1

Le premier quartile est la valeur qui est telle qu'il y ait au moins 25 % de valeurs inférieures ou égales.

Pour la déterminer le premier quartile :

➡ on range les valeurs par ordre croissant

➡ nous calculons $\frac{N}{4}$

➡ le premier quartile est l'arrondi par excès de $\frac{N}{4}$

3.6 Le troisième quartile Q_3

Le troisième quartile est la valeur qui est telle qu'il y ait au moins 75 % de valeurs inférieures ou égales.

Pour déterminer le troisième quartile :

➡ on range les valeurs par ordre croissant

➡ nous calculons $\frac{3N}{4}$

➡ le troisième quartile est l'arrondi par excès de $\frac{3N}{4}$

3.7 L'écart-type σ

L'écart-type est un nombre qui exprime la dispersion des valeurs observées autour de la moyenne. Pour faciliter ce calcul, on ajoute une colonne à notre tableau.

$$\sigma = \sqrt{\frac{1}{N} \sum_i n_i (x_i - \bar{x})^2}$$