

Fiche n°8

**INTERPRETER, REPRESENTER et TRAITER
DES DONNEES STATISTIQUES****I. Introduction : de l'importance des statistiques et des probabilités ?**

Les **statistiques** représentent le domaine des mathématiques qui permet d'**étudier des données réelles** que l'on a obtenues :

- soit par un **recensement** (recueil complet de toutes les données, sur toute la population étudiée) : ceci permet une image précise de ce que l'on désire observer mais pose des problèmes techniques évident s'il y a un grand nombre de données ;
 - soit par l'étude d'**échantillons** (recueil de données que sur une partie seulement de la population à étudier : par exemple, sondages...) : ceci présente donc quelques possibilités d'erreur (incertitude) qu'il faut minimiser.
- A ne pas confondre avec les **probabilités** qui proposent des modèles **pour prévoir sans avoir expérimenté** le résultat d'expérience dans lequel intervient le **hasard**...

Quelques repères dans le temps...

- Les premiers relevés d'hommes et de bien ont eu lieu **vers 3000 ans avant J.-C. en Mésopotamie** ;
- **L'Égypte des pharaons** organisait régulièrement des recensements notamment pour les impôts ;
- **Tycho Brahe** (1546-1601), astronome danois, utilise la moyenne arithmétique pour réduire les erreurs d'observations ;
- Premières analyses de situation de probabilité vers le XVII^e siècle : **Pierre de Fermat, Blaise Pascal**... ;
- Au XVIII^e et XIX^e siècle se développe la **théorie des erreurs** ;
- Essor à partir XX^e siècle :
 - En statistiques, les **ordinateurs** permettent de **nombreuses simulations**...
 - En probabilités, développement de la théorie actuelles des probabilités...

Quelle utilité dans le monde contemporain ?

- **Pour trouver et décrire une relation** : en médecine, on établit le risque cardiovasculaire lié au tabac en étudiant le pourcentage de fumeurs chez les cardiaques et le pourcentage de cardiaques chez les fumeurs et les non-fumeurs ;
- **Prendre une décision** : l'amélioration annuelle des semences de céréales par croisements successifs, les contrôles de fabrication et de fiabilité dans l'industrie, d'efficacité d'un médicament, etc. sont très dépendants des tests statistiques ;
- **Prévoir et planifier** : de nombreuses statistiques économiques sont publiques (INSEE) et servent par exemple de base aux négociations syndicales ou intergouvernementales.

Exemples de questions d'actualités :

- **Climat** : assiste-t-on à un réchauffement de la planète ?
- **Santé** : faut-il encore vacciner les enfants contre la variole ?
- **Paris sportifs** : une activité à risques ?
- **Qualité industrielle** : comment faire pour être « sûr » que dans un lot de 1000 piles électriques vendues, il y en a au moins 995 qui fonctionnent correctement ?
- **Météo** : fera-t-il beau dimanche ?
- **Population** : quel pays aura le plus d'habitant en 2020 ?

- **Jeux** : un dé qui affiche 241 fois la face « 6 » sur 1 000 lancers est-il truqué ?

II. Graphiques : comment choisir une bonne représentation ?

On peut représenter des données avec plusieurs graphiques selon :

- ce que l'on souhaite observer : le **caractère** ;
- ce que l'on souhaite montrer grâce au graphique : pour comparer des **effectifs**, pour répartir des **fréquences** ou pour **regrouper des données** (histogramme).

Quel type de **caractère** dois-je étudier et observer ?

Caractères numériques

(nombres ordonnés > axe des abscisses)

Exemples : âge, temps ou durée, masse, notes d'un contrôle, prix ou salaires, etc.

Caractères qualitatifs

Exemples : couleur des yeux, boisson préférée, langue parlée, secteurs d'activités, jours de la semaine, etc.

Pour comparer, trouver un maximum...

Diagramme en barres (bâtons)

La contribution aux émissions de CO2 par secteurs, en %, France, 2005

Pour comparer, trouver un maximum...

Diagramme en barres (bâtons)

Proportion des femmes parmi les députés

Pour observer une évolution.

Courbe

Morts pendant la seconde guerre mondiale

Pour voir une répartition (proportions)

Diagramme circulaire

Pour étudier des données regroupées.

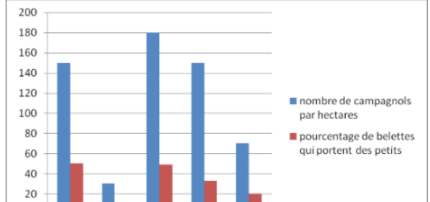
Histogramme (regroupement)

Notes par discipline (Pronote)

Pour comparer à une moyenne.

Et bien d'autres diagrammes...

Remarque On peut parfois comparer deux caractères sur un même graphique. Par exemple, ce graphique ci-contre permet de



comparer des populations entre proies (campagnols) et prédateurs (belettes).

III. Effectifs et fréquences

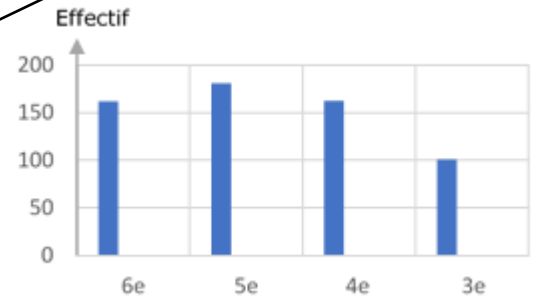
1. Effectifs

Le tableau ci-contre donne la répartition des **effectifs** des élèves dans un collège dont l'**effectif total** est de 607 élèves.

	6 ^{ème}	5 ^{ème}	4 ^{ème}	3 ^{ème}	Total
Effectifs	162	181	163	101	607

Exemple L'effectif des 5^{èmes} est de 181 élèves.

Graphique On peut représenter ces effectifs par un diagramme en barres où la hauteur de chaque barre est proportionnelle à l'effectif qu'elle représente.



2. Fréquences (et pourcentages)

Définition
$$\text{fréquence} = \frac{\text{effectif}}{\text{effectif total}}$$

	6 ^{ème}	5 ^{ème}	4 ^{ème}	3 ^{ème}	Total
Effectifs	162	181	163	101	607
Fréquences	0,267	0,298	0,268	0,167	1

Exemple

La fréquence des élèves de 3^{ème} dans ce collège est 0,167 environ car $\frac{101}{607} \approx 0,167$

Remarques ✕ Une fréquence est un nombre compris entre 0 et 1.

✕
$$\text{fréquence en \%} = \frac{\text{effectif}}{\text{effectif total}} \times 100$$

On obtient alors :

	6 ^{ème}	5 ^{ème}	4 ^{ème}	3 ^{ème}	Total
Effectifs	162	181	163	101	607
Fréquences (en %)	26,7	29,8	26,8	16,7	100

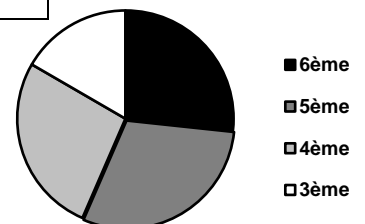
Exemple

Le pourcentage de 4^{ème} dans ce collège est 26,8 % environ car $\frac{163}{607} \times 100 \approx 26,8$

Graphique On peut représenter ces effectifs ou fréquences par un diagramme circulaire où la mesure de chaque angle est proportionnelle à l'effectif ou la fréquence qu'il représente.

$$\text{Mesure d'un angle (en degrés)} = \frac{\text{effectif}}{\text{effectif total}} \times 360$$

	6 ^{ème}	5 ^{ème}	4 ^{ème}	3 ^{ème}	Total
Effectifs	162	181	163	101	607
Mesures (en degrés)	96	107	97	60	360



Exemple L'angle représentant les 5^{èmes} mesure 107° environ car $\frac{181}{607} \times 360 \approx 107$

IV. Caractéristiques d'une série statistique : moyenne, médiane et étendue

1. Moyenne pondérée

EXERCICE TYPE 1 Déterminer la taille moyenne pour les 10 personnes suivantes

Taille (en m)	1,70	1,75	1,80	1,85	Total
Effectif	3	4	2	1	10

Solution $1,70 \times 3 + 1,75 \times 4 + 1,80 \times 2 + 1,85 \times 1 = 17,55$

$$17,55 \div 10 = 1,755$$

La taille moyenne de ces 10 personnes est **environ 1,76 m**.



EXERCICE TYPE 2 Déterminer la taille moyenne des élèves de cette classe.

Taille (en m)	[1,50 ; 1,60[[1,60 ; 1,70[[1,70 ; 1,80[[1,80 ; 2[Total
Centre	$(1,50 + 1,60) \div 2$ = 1,55	$(1,60 + 1,70) \div 2$ = 1,65	$(1,70 + 1,80) \div 2$ = 1,75	$(1,80 + 2) \div 2$ = 1,90	
Effectif	3	13	8	2	26

Solution Pour des données regroupées, on utilise le centre de l'intervalle pour calculer la moyenne (voir les calculs dans le tableau)

$$1,55 \times 3 + 1,65 \times 13 + 1,75 \times 8 + 1,90 \times 2 = 43,9$$

$$43,9 \div 26 \approx 1,69$$

La taille moyenne des élèves de la classe est **environ 1,69 m**.

2. Médiane d'une série statistique

Définition La **médiane** d'une série ordonnée est **une valeur** telle qu'il y ait **autant de valeurs inférieures** ou égales **que de valeurs supérieures** ou égales.

EXERCICE TYPE 3 Déterminer les médianes des séries de notes suivantes.

- série A : 13, 13, 20, 19, 18, 15, 15
- série B : 8, 8, 9, 12, 15, 17, 12, 11, 14, 14
- série C : 17, 14, 3, 16, 5, 17

Solution Pour déterminer une médiane, **il faut d'abord ordonner la série.**

- série A : $13 \leq 13 \leq 15 \leq \boxed{15} \leq 18 \leq 19 \leq 20$. La médiane de cette série est **15**.
 $\xleftarrow{3 \text{ notes}} \quad \quad \quad \xrightarrow{3 \text{ notes}}$

- série B : $8 \leq 8 \leq 9 \leq 11 \leq \boxed{12 \leq 12} \leq 14 \leq 14 \leq 15 \leq 17$. La médiane est **12**.
 $\xleftarrow{5 \text{ notes}} \quad \quad \quad \xrightarrow{5 \text{ notes}}$

- série C : $3 \leq 5 \leq \boxed{14 \leq 16} \leq 17 \leq 17$.
 $\xleftarrow{3 \text{ notes}} \quad \quad \quad \xrightarrow{3 \text{ notes}}$

La médiane de cette série est entre 14 et 16. Par habitude, on prendra alors la valeur centrale : la médiane de cette série C est donc **15**.

- Remarques
- ✕ Deux séries peuvent avoir la même moyenne mais pas la même médiane (séries B et C).
 - ✕ Deux séries peuvent avoir la même médiane mais pas la même moyenne (séries A et C).

	Série A	Série B	Série C
Médiane	15	12	15
Moyenne	≈ 16,1	12	12

3. L'étendue

Définition L'**étendue** d'une série est la différence entre les deux valeurs extrêmes.

EXERCICE TYPE 4 Déterminer l'étendue des séries A, B et C de l'exercice type 3.

- Solution
- série A : $20 - 13 = 7$. L'étendue de cette série est 7.
 - série B : $17 - 8 = 9$. L'étendue de cette série est 9.
 - série C : $17 - 3 = 14$. L'étendue de cette série est 14.

- Remarque
- ✕ La moyenne et la médiane donnent des informations sur la **position** et la répartition des valeurs.
 - ✕ L'étendue donne uniquement des informations sur la **dispersion**, mais pas sur la répartition entre la valeur minimale et la valeur maximale.