

Statistiques

« Selon les statistiques, il y a une personne sur cinq qui est déséquilibrée.

S'il y a 4 personnes autour de toi et qu'elles te semblent normales, c'est pas bon. »

Jean Claude Van Damme

Introduction

Les statistiques sont un domaine des mathématiques au développement assez récent mais dont les applications sont nombreuses et variées : leur but est de déduire des lois et des comportements généraux (de façon la plus fiable possible) à partir de l'étude d'un nombre limité d'exemples (un échantillon).

On les rencontre bien sûr dans l'étude des comportements humains (sondages, ...) et économiques, mais aussi dans l'industrie (fiabilité d'une machine, ...), en biologie....

Une étude statistique comporte généralement quatre étapes :

- Le recueil des données : sondages, recensements, enquêtes, mesures, ...
- La présentation des résultats : tableaux, graphiques, ...
- Le calcul de paramètres statistiques : moyenne, médiane, ...
- L'exploitation des données : savoir tirer des conclusions des calculs précédents.

Les statistiques fonctionnent avec un vocabulaire spécifique.

I. Vocabulaires des statistiques (rappels)

Définitions

Une étude statistique s'effectue sur un ensemble appelé **population**.

Les éléments de la population sont des **individus**.

Une étude statistique consiste à observer, étudier un aspect sur chaque individu nommé **caractère**.

Le caractère a différentes **valeurs** (ou modalités).

Il existe deux types de caractère :

- Les **caractères qualitatifs** : Les valeurs de la série ne sont pas des nombres. C'est le cas ici : « football », « basket », « tennis », « handball » et « danse ».
- Les **caractères quantitatifs** que l'on peut mesurer.

A. L'effectif

Définitions

L'**effectif** de chaque valeur est le nombre d'individus qui la composent.

L'**effectif total** est la somme des effectifs de toutes les valeurs.

Remarque

L'**effectif cumulé croissant** d'une valeur d'une série statistique est le nombre de termes de la série dont la valeur est inférieure ou égale à cette valeur.

L'**effectif cumulé décroissant** d'une valeur d'une série statistique est le nombre de termes de la série dont la valeur est supérieure ou égale à cette valeur.

B. La fréquence

Définition

La **fréquence** d'une valeur d'une série statistique est le quotient de l'effectif de cette valeur par l'effectif total.

Remarques

- Les fréquences sont toujours inférieures à 1 et la somme des fréquences est égale à 1.
- Les fréquences sont souvent exprimées en pourcentage (il suffit de multiplier la fréquence par 100).
- Dans des documents on peut trouver la fréquence cumulée croissante (ou décroissante) d'une valeur d'une série statistique. Elle est égale au quotient de l'effectif cumulé croissant (ou décroissant) correspondant par l'effectif total.

II. Paramètres d'une série statistique

A. Paramètres de position

1. La moyenne

Définition

La **moyenne** d'une série statistique est le quotient de la somme de toutes les valeurs de cette série par l'effectif total. On la note \bar{m} .

Exemple 1

Marie a indiqué dans le tableau suivant le temps qu'elle a passé devant la télévision la semaine dernière.

Jour	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Temps (en min)	61	53	114	60	42	126	132

Le temps quotidien moyen est :

$$\bar{m} = \frac{61+53+114+60+42+126+132}{7} = \frac{588}{7} = 84 \text{ (en min)}$$

Définition

La **moyenne pondérée** d'une série de données est la somme des produits des valeurs par leur effectif, divisée par l'effectif total.

Exemple 2

La taille en mètre de 100 requins blancs.

Taille (m)	1,5	2	2,5	3	3,5	4	4,5
Effectif	8	10	25	32	19	4	2

La taille moyenne est :

$$\bar{m} = \frac{1,5 \times 8 + 2 \times 10 + 2,5 \times 25 + 3 \times 32 + 3,5 \times 19 + 4 \times 4 + 4,5 \times 2}{100} \approx 2,82 \text{ (en m)}$$

Exemple 3

Un supermarché a relevé les dépenses (en €) de ses clients en 2h un jour donné.

Dépenses (en €)	[0 ; 30[[30 ; 60[[60 ; 100[[100 ; 120[
Effectif	12	25	42	67
Milieu de classe	15	45	80	110

Pour déterminer la moyenne, on détermine les milieux des classes de la série puis on effectue le calcul.

$$\bar{m} = \frac{15 \times 12 + 45 \times 25 + 80 \times 42 + 110 \times 67}{146} \approx 82,43 \text{ (en €)}$$

2. La médiane

Définition

La médiane d'une série statistique est une valeur du caractère qui partage la série en deux groupes (l'un des valeurs inférieures à la médiane, l'autre des valeurs supérieures) de même effectif.

Autrement dit

Cela signifie que 50% de la population a une valeur du caractère inférieure à la médiane, et que 50% de la population à une valeur du caractère supérieure à cette médiane.

Méthode

On range d'abord les valeurs du caractère dans l'ordre croissant et ensuite :

- Soit l'effectif total N est impair
La médiane est la valeur centrale c'est-à-dire la valeur de rang $\frac{N+1}{2}$.
- Soit l'effectif total N est pair
La médiane est la demi-somme (la moyenne) des données de rang $\frac{N}{2}$ et $\frac{N}{2} + 1$.

Exemple 1

Un boulanger teste les masses (en g) de 30 baguettes.

Masse (g)	235	237	238	239	240	241	243	245	247	249	250	251	253	255	257	260
Effectif	2	1	2	3	1	2	1	1	2	1	5	2	2	3	1	1

$N = 30$ L'effectif total est pair

La médiane est la demi-somme des données de rang 15 et 16 soit $m = \frac{247+249}{2} = 248$ (en g).

Exemple 2

la durée (en min) de connexion internet par jour de 43 familles.

Durée (min)	40	60	80	120	180	200	240	300
Effectif	2	9	11	7	5	2	4	3

$N = 43$ L'effectif total est impair

La médiane est la valeur de rang 22 soit $m = 80$ (en min).

B. Paramètres de dispersion

Définition

L'étendue d'une série statistique est égale à la différence entre la plus grande et la plus petite des données de la série.

Remarque

L'étendue est une mesure de **dispersion** des valeurs : plus l'étendue est grande, plus les valeurs sont dispersées.

Exemple 1

masse des baguettes.

$$e = 260 - 235 = 25 \text{ (en g)}$$

Exemple 2

connexion internet.

$$e = 300 - 40 = 260 \text{ (en min)}$$