Chapitre B2
Statistiques et probabilités

I. Définitions

La **statistique** est la science qui recueille et analyse mathématiquement des données numériques.
En statistique, on utilise deux types d'outils :

- Les **outils graphiques** (diagramme circulaire, diagramme bâtons, histogramme,…) ;
- Les **outils numériques** (moyenne, médiane, étendue,…)

Les statistiques étudient une **population** composée d'**individus** (objets, personnes, …) éventuellement classés suivant un **caractère** (couleur, taille, âge, …). Un caractère peut être **qualitatif** (sexe, couleur) ou **quantitatif** (taille, prix). Le nombre d'individus d'une classe de caractère est l'**effectif** de cette classe. Le nombre d'individus d'une population est l'**effectif total**.

Exemple : une usine produit des voitures. On a recensé les voitures produites en une journée.

Couleur de la voiture	Grise	Noire	Bleue	Rouge
Nombre de voitures produites en un jour	420	300	160	120

La **population** est l'ensemble des voitures produites en une journée.
Un **individu** est une voiture.
Le **caractère** étudié est la couleur.
L'**effectif** des voitures grises est 420.
L'effectif total est $420 + 300 + 160 + 120 = 1000$.
La **fréquence** d'une classe est le quotient de l'effectif de cette classe par l'effectif total.

$$(fréquence) = \frac{(effectif\ de\ la\ classe)}{(effectif\ total)}$$

Exemple

Dans la situation précédente, la fréquence de voitures grises égale $\frac{420}{1000}$.

Remarque

On peut exprimer cette fréquence en pourcentage.

$$\frac{420}{1000} = 0,42 = \frac{42}{100}$$

42% des voitures sont grises.

Exercice

Une usine produit des baskets. Voici un relevé de la production sur 10 minutes

Pointure	40	41	42	43	44	45
Nombre de chaussures produites	120	450	240	410	100	80

Quelle est la population étudiée ? Quels en sont les individus ? Quel est le caractère étudié ?
Calcule la fréquence des chaussures en 44.

Solution

La population est l'ensemble des baskets produites en 10 min.
Chaque basket est un individu. Le caractère étudié est la pointure.

$$\frac{100}{120+450+240+410+100+80} = \frac{100}{1400} \approx 0,071 \approx \frac{7,1}{100}$$

7,1% des chaussures produites sont en pointure 44.

II. Outils numériques

Si le caractère étudié dans une population est **quantitatif**, on peut calculer la valeur moyenne de ce caractère.

La **moyenne** tient compte des effectifs de chaque classe de caractère. Si n_1, n_2,… désignent les valeurs du caractère et si m_1, m_2,… désignent l'effectif des classes associées, alors la moyenne m de cette série statistique égale :

$$m = \frac{m_1 \times n_1 + m_2 \times n_2 + \ldots}{m_1 + m_2 + \ldots}$$

Exemple

On peut calculer la pointure moyenne des chaussures produites en 10 min dans une usine :

Pointure	40	41	42	43	44	45
Nombre de chaussures produites	120	450	240	410	100	80

La pointure moyenne égale :

$$m = \frac{40\times120+41\times450+42\times240+43\times410+44\times100+45\times80}{120+450+240+410+100+80} = \frac{58\,960}{1400} \approx 42$$

Remarque

Dans un devoir, si le calcul est trop long, on peut utiliser des points de suspension :

$$m = \frac{40\times120+\ldots+45\times80}{120+\ldots+80} = \frac{58\,960}{1\,400} \approx 42$$

L'**étendue** d'une série statistique est la **différence entre la plus grande valeur et la plus petite valeur** (de caractère) de cette série statistique.

Dans l'exemple précédent, l'étendue de la série égale $45 - 40 = 5$.

La **médiane** d'une série statistique est la valeur du caractère pour laquelle **la moitié ou moins de l'effectif total est de part et d'autre de cette valeur**.

Par exemple, on peut calculer la pointure médiane des chaussures produites dans l'usine.

1400 chaussures ont été produites, on cherche donc la pointure pour laquelle il y a au plus 700 chaussures de pointure supérieure et 700 chaussures de pointure inférieure.

La pointure médiane des chaussures produites est 42 car il y a 570 chaussures de pointure plus petite et 590 de pointure plus grande.

Exemples

- On dispose de la série suivante :

5 – 7 – 9 – 10 – 14 -17 – 18

La médiane de cette série est 10 car il y a moins de la moitié des valeurs de part et d'autre de 10.

- On dispose de la série suivante :

3 – 4 – 6 – 8 – 9 -11 – 15 – 18

La médiane de cette série est comprise entre 8 et 9 : on peut prendre 8 – 8,1 - … - 8,5 – 8,88 - …. – 9.

Dans certains livres, la médiane est systématiquement la **moyenne des valeurs extrêmes**, ici 8,5.

- On dispose de la série suivante :

4 – 4 – 4 – 5 – 9 – 9 – 11 – 13 – 14 – 17

10: 2 = 5. La médiane de cette série est comprise entre la 5ème et la 6ème valeur, soit 9.

- On dispose de la série suivante :

4 – 4 – 4 – 5 – 9 – 9 – 11 – 13 – 14

9: 2 = 4, 5. La médiane de cette série est la 5ème valeur, soit 9.

Exemple type

On a relevé les notes des élèves de la classe de 4D :

15 12 9 17 20 8 13 14 19 18 19 5 16 14 13
9 11 15 17

1. *Quelle est la population étudiée ? Quel est le caractère étudié ?*

2. *Calcule la moyenne des résultats.*

3. *Calcule l'étendue de cette série de notes.*

4. *Calcule la médiane.*

Solution

1. La population étudiée est l'ensemble des élèves de 4°D. Le caractère étudié est la note.

2. La moyenne des résultats est de :

$$M = \frac{15+\ldots+17}{19} = \frac{264}{19} \approx 14$$

3. L'étendue de cette série de notes est de $20 - 5 = 15$.

4. Je calcule la médiane. Je range les valeurs dans l'ordre croissant :

5 8 9 9 11 12 13 13 14 14 15 15 16 17
17 18 19 19 20

$\frac{19}{2} = 9,5$ La médiane est donc la 10ème valeur, soit 14.

Remarques

Pour calculer une médiane, on range les valeurs **dans l'ordre croissant**.

Si le nombre de valeurs est **impair**, il n'y a qu'**une seule valeur** pour la médiane. Si le nombre de valeurs est **pair**, il peut y avoir plusieurs valeurs pour la médiane.

Effectifs cumulés croissants et médiane

On dispose de la série statistique suivante :

Pointure	**40**	**41**	**42**	**43**	**44**	**45**
Effectif	120	450	240	410	100	80

Comment calcule-t-on la médiane de cette série ?

On utilise les effectifs cumulés croissants.

Pointure	**40**	**41**	**42**	**43**	**44**	**45**
Effectif	120	450	240	410	100	80
Effectifs cumulés croissants	120	570	810	1220	1320	1400

1220 est le nombre de chaussures dont la pointure est inférieure ou égale à 43. 1400 est l'effectif total. Comme il y a eu 1400 chaussures produites, la médiane est une valeur comprise entre la 700ème et la 701ème valeur.

La 1ère valeur jusqu'à la 120ème est 40. La 121ème valeur jusqu'à la 570ème valeur est 41. La 571ème valeur jusqu'à la 810ème valeur est 42. La 700ème et la 701ème valeur sont donc 42 : la médiane est donc 42.

Exemple de regroupement par classe

On relève la taille en cm des élèves de la classe de 4°1 :

160 – 140 – 165 – 150 – 138 – 158 – 170 – 160 – 155 – 160 – 140 – 150 – 170 – 155 – 155 – 129 – 160 – 163

On regroupe ces tailles en plusieurs classes et dans le tableau suivant :

Taille	[120 ;130[	[130 ;140[	[140 ;150[	[150 ;160[	[160 ;170[	[170 ;180[
Effectif	1	1	2	6	6	2

[120 ;130[est un intervalle qui rassemble tous les élèves dont la taille varie entre 120 cm et 130 cm. Les crochets indiquent que les élèves qui font exactement 120 cm sont dans cette classe et ceux qui font exactement 130 cm ne sont pas dans cette classe.

La fréquence des élèves dont la taille est comprise entre 140 et 150 cm est de $\frac{2}{18} = \frac{1}{9} \approx 0,11 \approx \frac{11}{100}$.

Avec ce tableau, on peut calculer une moyenne approchée à l'aide du centre des classes et la classe médiane à l'aide des effectifs cumulés croissants (ECC).

Taille	[120 ;130[	[130 ;140[	[140 ;150[	[150 ;160[	[160 ;170[	[170 ;180[
Centre	$\frac{120+130}{2} = 125$	135	145	155	165	175
Effectif	1	1	2	6	6	2
ECC	1	$1 + 1 = 2$	4	10	16	18

La taille moyenne est de

$$\frac{125\times1+135\times1+\ldots+175\times2}{1+1+\ldots+2} = \frac{2820}{18} \approx 157\ cm$$

Remarque

Ceci est une valeur approchée de la moyenne.

$$\frac{160+140+\ldots+163}{18} = \frac{2778}{18} \approx 154\ cm$$

Elle peut être légèrement différente de la valeur réelle.

$18 : 2 = 9$. La médiane est donc comprise entre la 9ème et la 10ème valeur. La classe médiane est donc [150 ;160[.

Dans cette classe, les valeurs sont 150-150-155-155-155-158.

155 est la 9ème valeur, 158 la 10ème valeur. La médiane est donc de 156,5 cm.

III. Probabilités

1. Activité préparatoire

Partie A

Lance 40 fois de suite une pièce de monnaie et note tous les résultats obtenus (P pour pile et F pour Face) :

Résultat	P	F	Total
Effectif	22	18	40
Fréquence	22/40	18/40	1

Partie B

Élève	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total	Fréquence
P	33	17	18	10	24	18	27	6	21	22	22	17	19	21	**275**	**0,49**
F	7	23	22	30	16	22	13	34	19	18	18	23	21	19	**285**	**0,51**

On remarque que, plus le nombre de lancers est important, plus la fréquence d'obtenir pile ou face se rapproche de 50%. Cette fréquence théorique définit la probabilité d'obtenir pile ou face dans cette expérience. On note P(F) la probabilité d'obtenir face et P(P) la probabilité d'obtenir pile.

$$P(F) = P(P) = \frac{1}{2}$$

2. Définitions

En probabilité, on définit une **expérience aléatoire** qui comporte une ou plusieurs **issues**.
Par exemple, on peut lancer un dé à 6 faces et regarder le résultat indiqué sur la face supérieure. L'expérience comporte 6 **issues** : tirer 1, 2, 3, 4, 5 ou 6.
Pour chaque expérience, on peut définir des **événements** réalisés par une, plusieurs ou même aucune issue.
« Tirer un nombre pair », « tirer un nombre », « ne tirer aucun nombre » sont des événements liés à l'expérience précédente.
Un événement réalisé par **aucune issue** est un **événement impossible.**
Un événement réalisé par **une seule issue** est un **événement élémentaire**.
Un événement réalisé par **toutes les issues** est un **événement certain**.
Si toutes les issues ont la même chance de se produire, on est dans une **situation d'équiprobabilité**.
Par exemple, le lancer d'un dé à six faces (non truqué) est une situation d'équiprobabilité : on a autant de chance de faire 1, 2, 3, 4, 5 ou 6.
Si on regarde la somme des résultats du lancer de 2 dés à six faces, ce n'est pas une situation d'équiprobabilité : on a plus de chances d'obtenir 7 que 2.
La **probabilité d'un événement** est le **quotient** entre le nombre d'occurrences où l'événement est réalisé et le nombre d'occurrences totales.

Conséquence

La **probabilité** d'un événement est un **nombre compris entre 0 et 1**.
La probabilité d'un événement **impossible** est **0**.
La probabilité d'un événement **certain** est **1**.
Par exemple, « obtenir un nombre pair » avec un dé à 6 faces a une probabilité de $\frac{3}{6}$.
Par exemple, « obtenir un valet » en tirant une carte dans un jeu de 52 cartes a une probabilité de $\frac{4}{52}$.
Deux événements sont **INCOMPATIBLES** s'ils ne peuvent pas se produire en même temps.
Deux événements sont **CONTRAIRES** s'ils sont incompatibles et si la somme de leurs probabilités vaut 1.

Exemple 1

On lance un dé à six faces. Si on obtient entre 1 et 4, on a perdu. Si l'on obtient 5 ou 6, on gagne.
Cette expérience possède **2 issues** : on perd ou on gagne.
La probabilité de gagner est de $\frac{2}{6} = \frac{1}{3}$.
La probabilité de perdre est de $\frac{4}{6} = \frac{2}{3}$.
On déduit que ce n'est **pas une situation d'équiprobabilité** : la probabilité des deux issues est différente.

Exemple 2

On dispose d'une urne qui contient 4 boules rouges, 3 boules vertes et une boule bleue. On tire une boule dans l'urne et on regarde sa couleur.
Cette expérience comporte **3 issues** : tirer une boule rouge, une boule verte ou une boule bleue.
La probabilité de tirer une boule rouge est de $P(R) = \frac{4}{8} = \frac{1}{2}$
La probabilité de tirer une boule verte est de $P(V) = \frac{3}{8}$
La probabilité de tirer une boule bleue est de $P(B) = \frac{1}{8}$
Ce n'est **pas une situation d'équiprobabilité** car les issues n'ont pas la même probabilité de se produire.

Remarque

On peut exprimer les probabilités en pourcentage. Par exemple,
$P(V) = \frac{3}{8} = 0,375 = \frac{37,5}{100}$