

# CHAPITRE IV



## Statistiques

---

**T**<sup>OUS</sup> le reste

**Plan du cours :**

a

**Programme (BO n° 30 du 26-7-2018) :**

— a

**Compétences :**

— a



## SITUATION INITIALE : Le paradoxe de Simpson — Les calculs rénaux

Un laboratoire médical vient de mettre au point deux traitements pour éliminer les calculs rénaux. Ces traitements ont été testés sur deux groupes de 350 personnes. Il y a deux types de calculs rénaux : les petits calculs rénaux inférieurs à 2 mm et les gros calculs.

Voici les résultats des tests effectués avec ces deux traitements :

**Test du Arnaclam 300 mg**

	Réussite	Échec	Total
Petits calculs	81	6	
Gros calculs	192		
Total			

**Test du Arnodix 750 mg**

	Réussite	Échec	Total
Petits calculs		36	
Gros calculs	55		80
Total			

1. Compléter les deux tableaux de résultats en utilisant les données de l'énoncé.
2. Compléter le tableau suivant en calculant en pourcentage à l'unité près les taux de réussite.

	Taux de réussite Arnaclam 300 mg	Taux de réussite Arnodix 750 mg
Petits calculs rénaux		
Gros calculs rénaux		
Total		

- 3.a Quel est le traitement le plus efficace sur les petits calculs rénaux?
  - 3.b Quel est le traitement le plus efficace sur les gros calculs rénaux?
  - 3.c Quel est le traitement le plus efficace sur l'ensemble de tous les calculs rénaux?
4. Que pensez-vous de cette situation?

*La plus lointaine mention d'un cas analogue remonte à 1899, où le mathématicien anglais Karl Pearson décrit des données équivalentes. Plus tard en 1903, Undy Yule redécouvrit le phénomène et le Britannique Edward Simpson écrivit en 1951 un article où cette singularité statistique était soigneusement étudiée et discutée.*

*De nombreux cas réels présentent cette inversion de résultat lorsqu'on regroupe plusieurs catégories complémentaires en une seule. De nombreux cas en médecine ont été rapportés. Le paradoxe a aussi été rencontré en démographie, dans l'analyse de match de basket-ball, dans l'étude de risque d'accidents... [4] [1]*





## SITUATION INITIALE : Le paradoxe de Simpson — Cigarettes et taux de mortalité

En 1972, à Whickham, une ville du nord-est de l'Angleterre, un sondage a été effectué afin d'éclairer des travaux sur les maladies cardiaques (Tunbridge et al. 1977). Une suite de cette étude a été menée vingt ans plus tard (Vanderpump et al. 1995). Certains des résultats avaient trait au tabagisme et cherchaient à savoir si les individus étaient toujours en vie lors de la seconde étude. Par simplicité, nous nous restreindrons aux femmes et parmi celles-ci aux 1314 qui ont été catégorisées comme « fumant actuellement » ou « n'ayant jamais fumé ». La survie à 20 ans a été déterminée pour l'ensemble des femmes du premier sondage.

Voici quelques résultats :

18-34 ans	Fumeuses	Non-fumeuses	Total
En vie	174	213	387
Décédée	5	6	11
<b>Total</b>	179	219	398

35-50 ans	Fumeuses	Non-fumeuses	Total
En vie	159	145	304
Décédée	36	16	52
<b>Total</b>	195	161	356

51-64 ans	Fumeuses	Non-fumeuses	Total
En vie	103		
Décédée		43	
<b>Total</b>		159	318

64 ans et plus	Fumeuses	Non-fumeuses	Total
En vie		28	
Décédée	42		
<b>Total</b>	49		

1. Compléter les deux tableaux restants en tenant compte des informations fournies.
2. Pour chacune des quatre tranches d'âge, calculer le taux de mortalité des fumeuses et des non-fumeuses en pourcentages arrondis au dixième près.  
*Le taux de mortalité est le ratio entre le nombre de personnes décédées et le nombre total de personnes considérées.*
3. Que constatez-vous en comparant ces taux de mortalité pour chaque tranche d'âge.
4. En cumulant les données de ces quatre tableaux, déterminer le taux de mortalité des fumeuses et des non-fumeuses sur l'ensemble des 1313 femmes interrogées.
5. Complétez le tableau de synthèse suivant :

Taux de mortalité en pourcentage	Fumeuses	Non-fumeuses
<b>18-34 ans</b>		
<b>35-50 ans</b>		
<b>50-64 ans</b>		
<b>64 ans et plus</b>		
<b>Ensemble</b>		

6. Que remarquez-vous? Comment pouvez-vous expliquer ce résultat?

*La plus lointaine mention d'un cas analogue remonte à 1899, où le mathématicien anglais Karl Pearson décrit des données équivalentes. Plus tard en 1903, Undy Yule redécouvrit le phénomène et le Britannique Edward Simpson écrivit en 1951 un article où cette singularité statistique était soigneusement étudiée et discutée.*

*De nombreux cas réels présentent cette inversion de résultat lorsqu'on regroupe plusieurs catégories complémentaires en une seule. De nombreux cas en médecine ont été rapportés. Le paradoxe a aussi été rencontré en démographie, dans l'analyse de match de basket-ball, dans l'étude de risque d'accidents... [4] [1]*







## SITUATION INITIALE : Les équipes de basket-ball

Voici la taille en centimètres des joueurs de deux équipes de basket-ball (même s'il y a cinq joueurs sur le terrain en même temps, il faut tenir compte des remplaçants) :

**Lakers de la Ramée** : 178 cm – 196 cm – 165 cm – 209 cm – 162 cm – 198 cm – 196 cm – 197 cm – 163 cm – 176 cm – 196 cm

**Celtics de Tibaous** : 183 cm – 185 cm – 181 cm – 187 cm – 196 cm – 183 cm – 176 cm – 188 cm – 207 cm – 185 cm – 165 cm

On souhaite comparer la taille des joueurs de ces deux équipes.

1. Calculer la moyenne des tailles en centimètres de chacune des deux équipes. Que pouvez-vous dire de ces résultats?
2. Déterminer la plus petite taille, la plus grande taille et l'écart entre la plus petite et la plus grande taille pour chacune des deux équipes. Que pouvez-vous dire de ces résultats?
3. Pour chacune de ces deux équipes, classer les tailles dans l'ordre croissant. Que pouvez-vous dire de ce classement?
4. Compléter le tableau suivant :

### Analyse des tailles des Lakers de La Ramée

Taille (cm)	[160; 170[	[170; 180[	[180; 190[	[190; 200[	[200; 210[	Total
Effectif						
Fréquence						

### Analyse des tailles des Celtics de Tibaous

Taille (cm)	[160; 170[	[170; 180[	[180; 190[	[190; 200[	[200; 210[	Total
Effectif						
Fréquence						

Que pouvez-vous en dire?

5. Voici les tailles de l'équipe des Hornets des Pradettes :

Taille (cm)	[160; 170[	[170; 180[	[180; 190[	[190; 200[	[200; 210[	Total
Effectif	3	2	1	2	3	
Fréquence						

Complétez ce tableau.

Pouvez-vous comparer cette équipe avec les deux précédentes?



---

## II — Annexes

---

### 1 Exercices

**EXERCICE N° 4.1 : Un exercice**



# STATISTIQUES

## VOCABULAIRE

Une **série statistique** est une liste de valeurs obtenues en étudiant une **population** (des élèves, des plantes, des factures...). Pour chaque **individu** de la population étudiée on peut observer un ou plusieurs **caractères** (tailles, masse, âge, prix, couleur...), c'est à dire une information. Un caractère peut être **qualitatif** (couleur, difficulté, goût...) ou **quantitatif** (quantité, nombre, prix...).

On connaît parfois toutes les valeurs d'une série statistiques. Quelquefois on ne connaît que la **répartition** des valeurs étudiées.

L'**effectif total** d'une série désigne le nombre total d'individu étudié. Dans un tableau de répartition on utilise le mot **effectif** pour le nombre d'individu concerné par une valeur du caractère.

La **fréquence** d'une valeur du caractère étudié correspond au quotient de l'effectif de ce caractère sur l'effectif total. Une fréquence peut s'exprimer sous forme d'une fraction, d'un pourcentage ou d'un nombre décimal approché ou non.

### EXEMPLES :

Voici une première série qualitative : la couleur des yeux de 10 personnes :

Bleu – Bleu – Vert – Vert – Vert – Marron – Marron – Marron – Marron – Noir

Voici une seconde série quantitative : les notes d'un groupe de 9 élèves au diplôme de fin d'année :

10 – 05 – 15 – 20 – 11 – 15 – 15 – 03 – 17

Voici une troisième série quantitative : la répartition des notes sur les 156 élèves de dernière année :

Notes	[0;5[	[5;10[	[10;15[	[15;20]
Effectif	26	54	60	16

## MOYENNE ARITHMÉTIQUE ET PONDÉRÉE

La **moyenne** ou **moyenne arithmétique** de la série de  $n$  valeurs :  $x_1, x_2, x_3, \dots, x_n$  est :

$$\frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

La **moyenne pondérée** de la série de  $n$  valeurs  $x_1, x_2, x_3, \dots, x_n$  pondérées par les nombres  $a_1, a_2, a_3, \dots, a_n$  est :

$$\frac{a_1 \times x_1 + a_2 \times x_2 + a_3 \times x_3 + \dots + a_n \times x_n}{a_1 + a_2 + a_3 + \dots + a_n}$$

La moyenne d'une série statistique est un nombre qui correspond à un partage équitable de toutes les valeurs de la série.

### EXEMPLES :

La première série est qualitative, la moyenne n'a pas de sens pour cette série.

La seconde série a pour moyenne :

$$\frac{10 + 5 + 15 + 20 + 11 + 15 + 15 + 3 + 17}{9} = \frac{111}{9} \approx 12,33 \text{ à } 0,01 \text{ près.}$$

Pour la troisième série, il faut calculer la moyenne des centres des intervalles pondérée par l'effectif.

$$\frac{2,5 \times 26 + 7,5 \times 54 + 12,5 \times 60 + 17,5 \times 16}{26 + 54 + 60 + 16} = \frac{1500}{156} \approx 9,62 \text{ à } 0,01 \text{ près.}$$

## ÉTENDUE

L'**étendue** d'une série statistique est l'écart entre la valeur maximale et minimale de la série.

L'étendue donne une information sur la dispersion des valeurs de la série : plus l'étendue est petite moins la série est dispersée.

### EXEMPLE :

L'étendue de la deuxième série est  $20 - 3 = 17$

Pour la deuxième série on peut seulement dire que l'étendue est inférieure ou égale à 20.

## MÉDIANE

La **médiane** d'une série statistique est une valeur du caractère qui partage la série en deux séries ayant le même effectif.

La moitié des valeurs sont inférieures à la médiane, l'autre moitié est supérieure.

La médiane donne une information sur la dispersion des valeurs de la série. Son écart avec la moyenne est souvent intéressant.

### MÉTHODE :

Pour calculer la médiane d'une série statistique il faut classer les valeurs du caractère dans l'ordre croissant puis déterminer la valeur centrale.

- si l'effectif est impair,  $2n + 1$ , la médiane est la  $n + 1^{\text{e}}$  valeur;
- si l'effectif est pair,  $2n$ , la médiane est la moyenne de la  $n^{\text{e}}$  et  $n + 1^{\text{e}}$  valeur.

### EXEMPLES :

Pour la deuxième série, l'effectif total est impair :  $9 = 2 \times 4 + 1$ , la médiane est la  $4 + 1 = 5^{\text{e}}$  valeur soit 15.

Pour la troisième série, l'effectif total est pair :  $156 = 2 \times 78$ , la médiane est la moyenne de la  $78^{\text{e}}$  et  $79^{\text{e}}$  valeurs.

D'après le tableau cette médiane se situe dans l'intervalle  $[5; 10[$ .

