Chapitre V

Statistiques

Sommaire

Infox

Le paradoxe de Simpson — Les calculs rénaux

Troisième

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 décrivit 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... [?] [?]

Infox

Le paradoxe de Simpson — Les calculs rénaux — Correction

✧✧

Infox

Le paradoxe de Simpson — Cigarettes et mortalité

Troisième

✧✧

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
Total	179	219	398

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

51-64 ans	Fumeuses	Non-fumeuses	**Total**
En vie	103		
Décédée		43	
Total		159	318

64 ans et plus	Fumeuses	Non-fumeuses	**Total**
En vie		28	
Décédée	42		
Total	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 1 314 femmes intérrogées.

5. Complétez le tableau de synthèse suivant :

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

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 décrivit 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...

Infox

Le paradoxe de Simpson — Cigarettes et mortalité — Correction

✧✧

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

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

51-64 ans	Fumeuses	Non-fumeuses	**Total**
En vie	103	116	219
Décédée	56	43	99
Total	159	159	318

64 ans et plus	Fumeuses	Non-fumeuses	**Total**
En vie	7	28	35
Décédée	42	165	207
Total	49	193	242

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.

Les 18-34 ans

Fumeuses : $\dfrac{5}{179} \approx 0,028$ soit $2,8\ \%$

Non-fumeuses : $\dfrac{6}{219} \approx 0,027$ soit $2,7\ \%$

Les 35-50 ans

Fumeuses : $\dfrac{36}{195} \approx 0,185$ soit $18,5\ \%$

Non-fumeuses : $\dfrac{16}{161} \approx 0,099$ soit $9,9\ \%$

Les 51-64 ans

Fumeuses : $\dfrac{56}{159} \approx 0,352$ soit $35,2\ \%$

Non-fumeuses : $\dfrac{43}{159} \approx 0,27$ soit $27\ \%$

Les 64 ans et plus

Fumeuses : $\dfrac{42}{49} \approx 0,857$ soit $85,7\ \%$

Non-fumeuses : $\dfrac{165}{193} \approx 0,855$ soit $85,5\ \%$

3. Que constatez-vous en comparant ces taux de mortalité pour chaque tranche d'âge.

Globalement le taux de mortalité est supérieur pour les fumeuses que les non-fumeuses.
Pour la population jeune ou très âgée les taux sont similaires pour des raisons simples à comprendre.

4. En cumulant les données de ces quatre tableaux, déterminer le taux de mortalité des fumeuses et des non-fumeuses sur l'ensemble des 1 313 femmes interrogées.

Total des fumeuses : $179 + 195 + 159 + 49 = 582$ et nombre de décès dans cette population : $5 + 36 + 56 + 42 = 139$.

Total des non-fumeuses : $219 + 161 + 159 + 193 = 732$ et nombre de décès dans cette population : $6 + 16 + 43 + 165 = 230$.

Le taux de mortalité chez les fumeuses : $\dfrac{139}{582} \approx 0,239$ soit $23,9\ \%$.

Le taux de mortalité chez les non-fumeuses : $\dfrac{230}{732} \approx 0,314$ soit $31,4\ \%$.

5. Complétez le tableau de synthèse suivant :

Taux de mortalité en pourcentage	Fumeuses	Non-fumeuses
18-34 ans	2,8 %	2,7 %
35-50 ans	18,5 %	9,9 %
50-64 ans	35,2 %	27 %
64 ans et plus	85,7 %	85,5 %
Ensemble	23,9 %	31,4 %

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

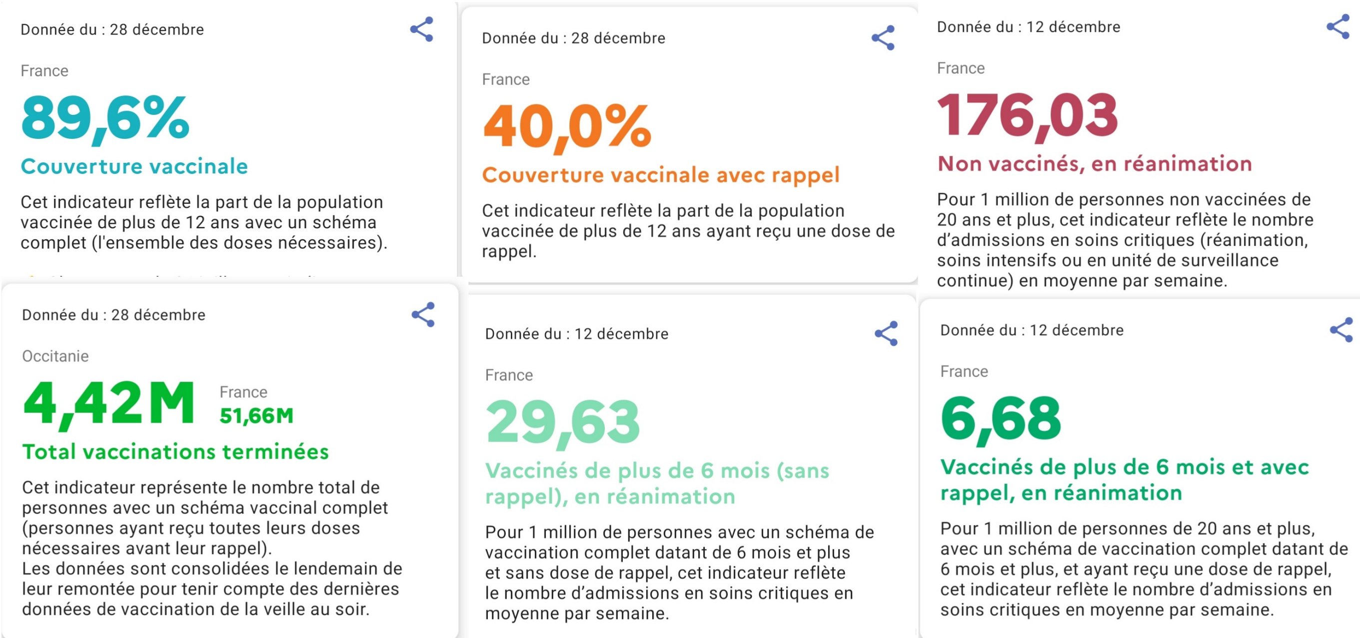
À rédiger !

Le paradoxe de Simpson — Covid et vaccin

Troisième

✧✧

Voici les chiffres fournis par l'application TousAntiCovid à la date du 29 décembre 2021 :



Sur les réseaux sociaux, certains Anti-vaccins prétendent « qu'il y a plus de vaccinés que de non vaccinés en réanimation aux urgences de l'hopital ». Cela prouverait que le vaccin est inefficace ce qui irait dans le sens de leur opinion.

On souhaite utiliser les informations de l'application TousAntiCovid pour vérifier cette information.

1. Déterminer l'effectif de la part de la population française de plus de 12 ans.

2. Déterminer le nombre de personnes non vaccinés et le nombre de ceux ayant reçu une dose de rappel.

3. Compléter le tableau suivant :

	Effectif total	Effectif en réanimation	Fréquence (%)
Vaccinés			
Dont ayant reçu une dose de rappel			
Non vaccinés			
Total			

4. Que peut-on déduire de ces informations et que peut-on répondre au sujet de cette information ?

Le paradoxe de Simpson — Covid et vaccin — Correction

Infox

Les données fournies par l'application TousAntiCovid mélangent des informations datant du 28 décembre 2021 et d'autres (les taux de malades en réanimation) qui datent du 12 décembre. Il y a donc un biais dans les calculs que nous allons faire. À cette date l'épidémie est dans une phase exponentielle. On peut donc imaginer que nos résultats vont être sous-évalué.

1. On constate que $89,6$ % de la population des plus de 12 ans correspond à $55,66$ M de personnes soit $55\,660\,000$.

Vaccinés	$89,6$	$55,66$ M
Total	100	$\dfrac{55,6\text{ M}\times 100}{89,6}\approx 62,05$ M

L'échantillon de la population concernée par ces statistiques correspond à $62,05$ M soit $62\,050\,000$ de personnes.

2. Comme $89,6$ % des personnes sont vaccinés, $100\,\% - 89,6\,\% = 10,4$ % ne le sont pas.

$\dfrac{10,4}{100}\times 62,05\text{ M}\approx 6,45$ M de personnes non vaccinés.

$\dfrac{40}{100}\times 62,05\text{ M} = 24,82$ M de personnes ayant une dose de rappel.

3. Compléter le tableau suivant :

	Effectif total	Effectif en réanimation	Fréquence (%)
Vaccinés	$51,66$ M	$51,66\text{ M}\times\dfrac{29,63}{1\text{ M}}\approx 1\,531$	$\dfrac{1\,531}{2\,667}\approx 0,57\approx 57$ %
Dont ayant reçu une dose de rappel	$24,82$ M	$24,82\text{ M}\times\dfrac{6,68}{1\text{ M}}\approx 166$	$\dfrac{166}{2\,667}\approx 0,06\approx 6$ %
Non vaccinés	$6,45$ M	$6,45\text{ M}\times$ ă $\dfrac{176,03}{1\text{ M}}\approx 1\,136$	$\dfrac{1\,136}{2\,667}\approx 0,43\approx 43$ %
Total	$62,05$ M	$1\,531+1\,136=2\,667$	100 %

4. On constate que les vaccinés sont majoritaires en réanimation. 57 % de vaccinés contre 43 % de non-vaccinés.
Cela semble indiquer que le vaccin manque d'efficacité.

Cependant quand on observe les fréquences qui concernent les non-vaccinés, les vaccinés et les doses de rappel, on constate que les non-vaccinés sont présents à $\dfrac{176,03}{1\,000\,000}$ dans les réanimations.

En comparant les fréquences on constate que :

— $176,03$ non vaccinés sur un million sont en réanimation ;

— $29,63$ vaccinés sur un million ;
— $6,68$ dose de rappel sur un million.

Comme $\dfrac{176,03}{29,63} \approx 5,94$: cela signifie qu'il est environ six fois plus probable d'aller en réanimation quand on est vacciné que quand on ne l'est pas !

Comme $\dfrac{176,03}{6,68} \approx 26,35$: il est environ vingt-six fois plus probable d'aller en réanimation quand on a une dose de rappel que quand on n'est pas vacciné.

Cela semble contredire l'étude des proportions de vaccinés et de non-vaccinés en réanimation !

Cette contradiction apparente est due au fait que la proportion de non vaccinés est faible dans la population. La différence de taille des échantillons entre les vaccinés et les non-vaccinés produit ce paradoxe de type paradoxe de Simpson.

En poussant ce raisonnement à l'extrême, quand 100 % de la population sera vaccinée alors 100 % des personnes en réanimation seront vaccinés (car le vaccin ne garantit pas une totale imunité...).

En effet, la majorité des personnes en réanimation sont des gens vaccinés. Cela ne prouve qu'une seule chose : qu'une majorité de la population est maintenant vaccinée !
a

Situation Initiale

Les équipes de basket-ball

Troisième

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 – 211 cm – 162 cm – 198 cm – 196 cm – 197 cm – 163 cm – 173 cm – 196 cm
Celtics de Tibaous : 185 cm – 185 cm – 179 cm – 187 cm – 196 cm – 183 cm – 176 cm – 188 cm – 206 cm – 184 cm – 166 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 :

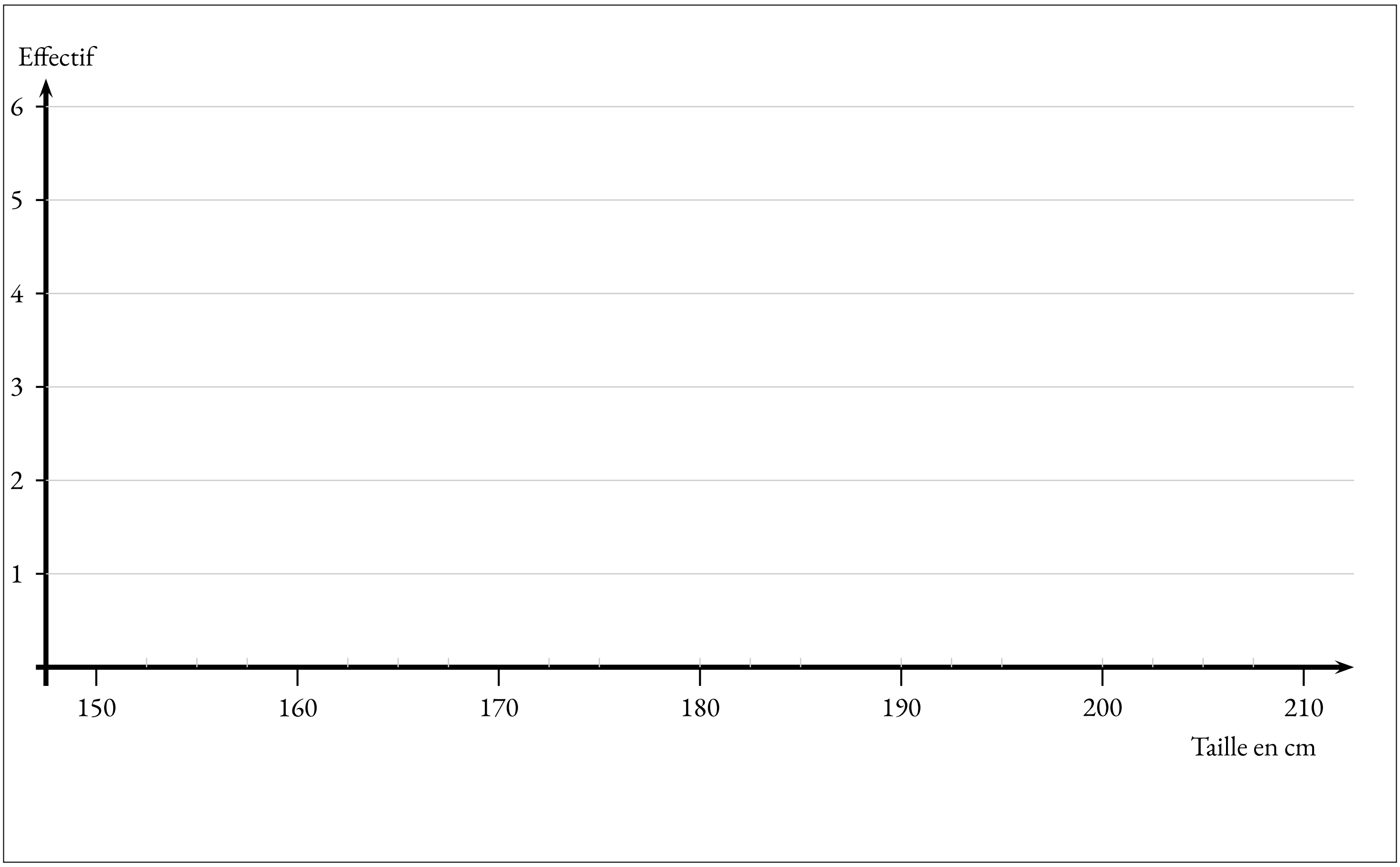
Taille (cm)	[160; 170[	[170; 180[	[180, 190[	[190; 200[	[200; 210[	Total
Effectif	3	2	1	2	3	
Fréquence						

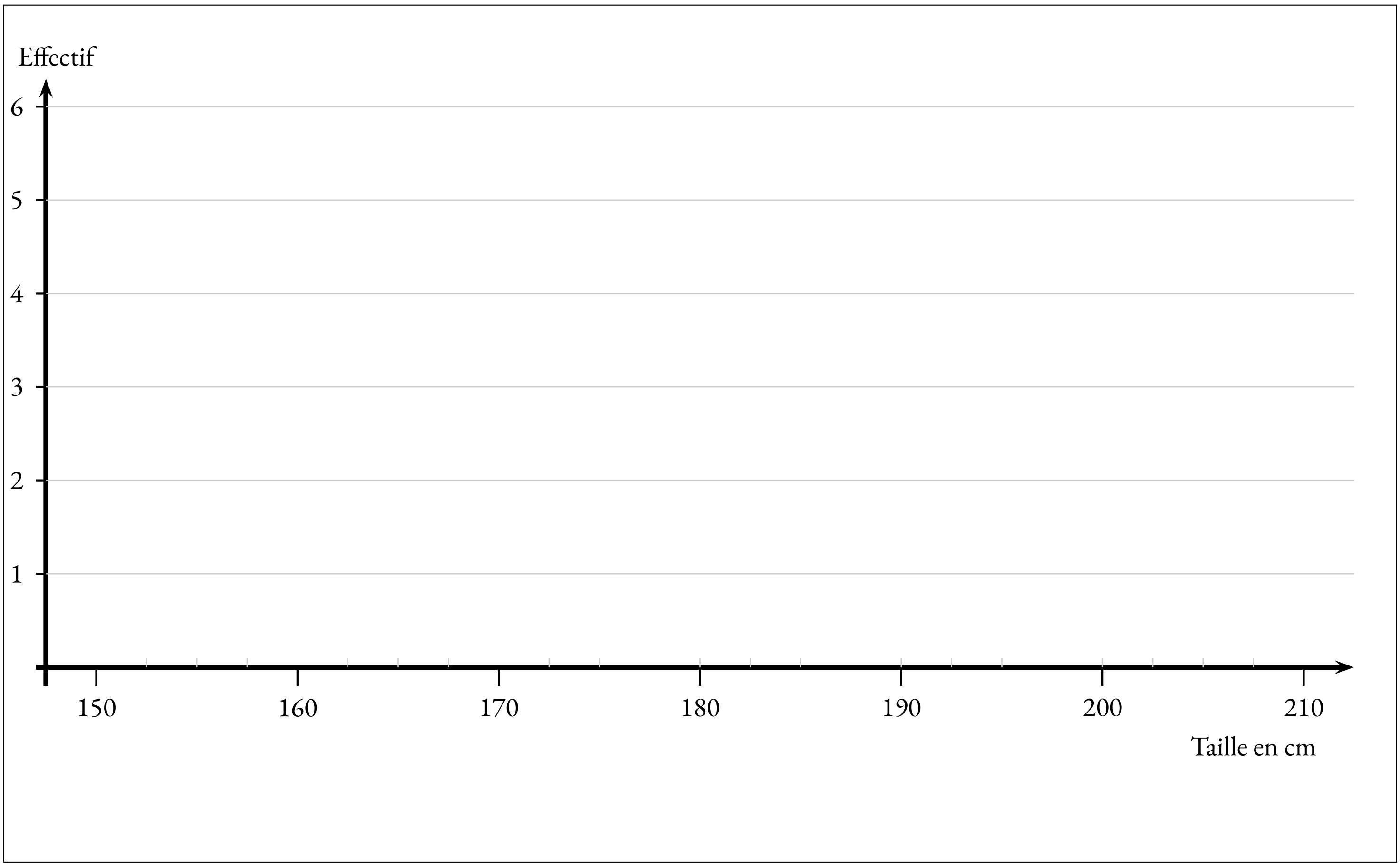
Complétez ce tableau.

6. Déterminer, avec ces informations, la moyenne, l'étendue et la médiane de la série des tailles des joueurs des Hornets.

7. Représenter graphiquement, sous forme de diagrammes en batons, les tailles des joueurs de chacunes de ces équipes.

Pouvez-vous comparer cette équipe avec les deux précédentes ?


Effectif
6
5
4
3
2
1
150
160
170
180
190
200
210
Taille en cm


Effectif
6
5
4
3
2
1
150
160
170
180
190
200
210
Taille en cm

Les équipes de basket-ball — Correction

✧✧

Situation Initiale

1. Moyen
des tailles des Lakers de la Ramée :

$$\frac{178\ cm + 196\ cm + 165\ cm + 211\ cm + 162\ cm + 198\ cm + 196\ cm + 197\ cm + 163\ cm + 173\ cm + 196\ cm}{11} = \frac{2\,035\ cm}{11} = 185\ cm$$

Moyenne des tailles des Celtics de Tibaous :

$$\frac{185\ cm + 185\ cm + 179\ cm + 187\ cm + 196\ cm + 183\ cm + 176\ cm + 188\ cm + 206\ cm + 184\ cm + 166\ cm}{11} = \frac{2\,035\ cm}{11} = 185\ cm$$

2. Le plus petit des Lakers mesure $163\ cm$, le plus grand mesure $211\ cm$.
L'écart entre le plus grand et le plus petit s'appelle **l'étendue** de la série statistique.
L'étendue est égale à $211\ cm - 163\ cm = 48\ cm$

Le plus petit des Celtics mesure $166\ cm$, le plus grand mesure $206\ cm$.
L'écart entre le plus grand et le plus petit s'appelle **l'étendue** de la série statistique.
L'étendue est égale à $206\ cm - 166\ cm = 40\ cm$

On constate que l'étendue pour la série des tailles des Lakers est supérieure à celle des Celtics. C'est un indicateur de **dispersion**. Cela signifie que les tailles des Lakers sont réparties sur un plus grand intervalle, elles sont moins regroupées que celles des Celtics.

3. On classe dans l'ordre croissant la tailles des joueurs :

Lakers : $162\ cm \leqslant 163\ cm \leqslant 165\ cm \leqslant 173\ cm \leqslant 178\ cm \leqslant 196\ cm \leqslant 196\ cm \leqslant 196\ cm \leqslant 197\ cm \leqslant 198\ cm \leqslant 211\ cm$

Celtics : $166\ cm \leqslant 176\ cm \leqslant 179\ cm \leqslant 183\ cm \leqslant 184\ cm \leqslant 185\ cm \leqslant 185\ cm\breve{a} \leqslant 187\ cm \leqslant 188\ cm \leqslant 196\ cm \leqslant 206\ cm$

En observant ce classement, on peut s'intéresser à la valeur centrale, celle qui partage l'effectif en deux.
Comme l'effectif total de ces deux séries est 11 et comme $11 = 5 + 1 + 5$, la sixième valeur de ce classement partage la série en deux séries d'effectif égaux. Cette valeur s'appelle **la médiane** de la série.

Pour les Lakers, la médiane vaut $196\ cm$. La moitié des Lakers mesurent au plus $196\ cm$, l'autre moitié mesure au moins $196\ cm$.
Pour les Celtics, la médiane vaut $185\ cm$. La moitié des Celtics mesurent au plus $185\ cm$, l'autre moitié mesure au moins $185\ cm$.

4.

Analyse des tailles des Lakers de La Ramée

Taille (cm)	$[160;170[$	$[170;180[$	$[180,190[$	$[190;200[$	$[200;210[$	Total
Effectif	3	2	0	5	1	11
Fréquence	$\frac{3}{11} \approx 27,3\ \%$	$\frac{2}{11} \approx 18,2\ \%$	$0\ \%$	$\frac{5}{11} \approx 45,5\ \%$	$\frac{1}{11} \approx 9,1\ \%$	$100\ \%$

Analyse des tailles des Celtics de Tibaous

Taille (cm)	$[160;170[$	$[170;180[$	$[180,190[$	$[190;200[$	$[200;210[$	Total
Effectif	1	2	6	1	1	11
Fréquence	$\frac{1}{11} \approx 9,1\ \%$	$\frac{2}{11} \approx 18,2\ \%$	$\frac{6}{11} \approx 54,5\ \%$	$\frac{1}{11} \approx 9,1\ \%$	$\frac{1}{11} \approx 9,1\ \%$	$100\ \%$

On constate à nouveau que les tailles des Celtics sont regroupées autour de la moyenne, dans l'intervalle $[180\ cm; 190\ cm[$.
Pour les Lakers, la dispersion est plus importante.

5.

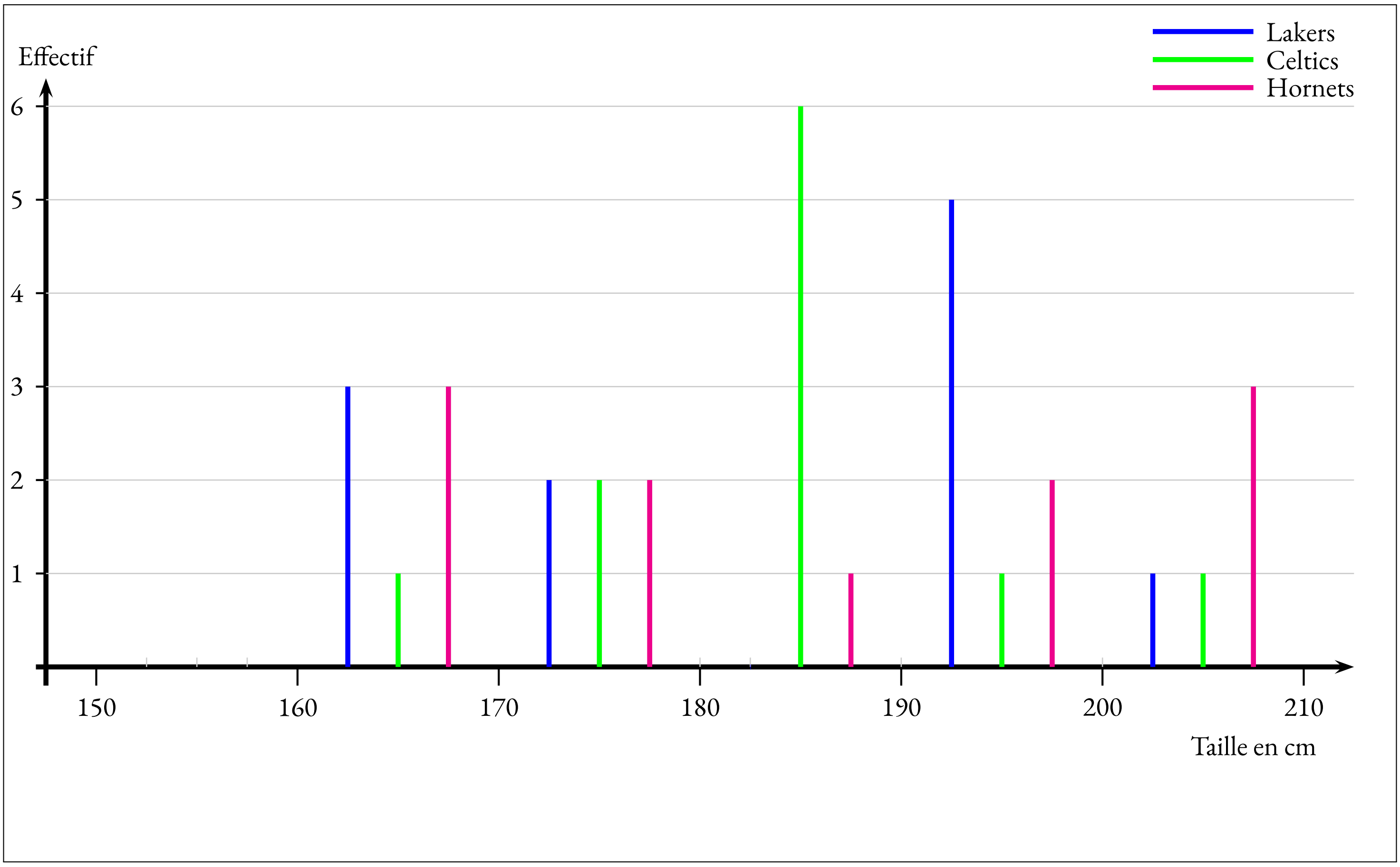
Taille (cm)	[160; 170[	[170; 180[	[180, 190[	[190; 200[	[200; 210[	Total
Effectif	3	2	1	2	3	11
Fréquence	$\frac{3}{11} \approx 27,3$ %	$\frac{2}{11} \approx 18,2$ %	$\frac{1}{11} \approx 9,1$ %	$\frac{2}{11} \approx 18,2$ %	$\frac{3}{11} \approx 27,3$ %	100 %

6. On ne peut pas obtenir ni la moyenne, ni la médiane, ni l'étendue de manière exacte puisque nous n'avons pas toute la série statistique. On peut cependant prendre les centres des intervalles pour effectuer ces calculs.

Le plus petit de cette équipe à une taille d'environ 165 cm, le centre de l'intervalle $[160\ cm; 170\ cm[$.
Le plus grand a une taille d'environ 205 cm.
L'étendue est vaut donc environ $205\ cm - 165\ cm = 40\ cm$.

Pour la moyenne on utilise la même méthode : $\dfrac{3 \times 165\ cm + 2 \times 175\ cm + 185\ cm + 2 \times 195\ cm + 3 \times 205\ cm}{11} = \dfrac{2\,035\ cm}{11} = 185\ cm$.

On observe aussi que la médiane est comprise entre 180 cm et 190 cm, soit environ 185 cm.



Les Hornets sont l'équipe dont les tailles sont les plus équilibrées.
Les Celtics ont des tailles très regroupées autour de la moyenne.
Les Lakers sont ceux dont la taille est la plus dispersée.

I — Moyenne arithmétique et pondérée

Évaluation

Troisième

Exercice n° 1

(6 points)

On pose :

$\mathbf{f(x) = 5x(3x-1) - 7(2x-3) - 8x^2}$ et $\mathbf{g(x) = (3x-7)(2x+3) + (3x-7)(4x-5)}$

1. Développer et réduire $f(x)$.

2. Développer et réduire $g(x)$.

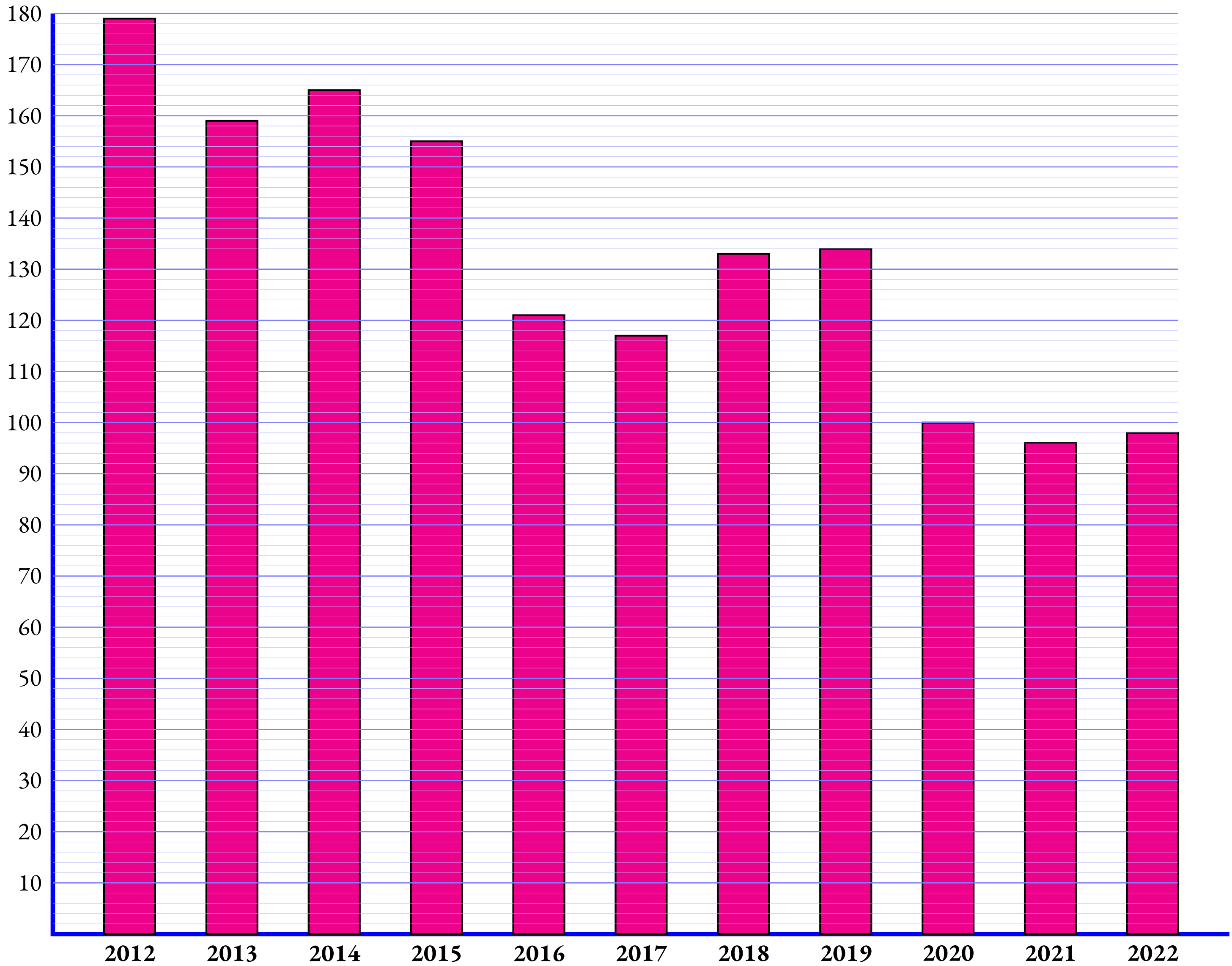
3. Factoriser $g(x)$.

Exercice n° 2

(8 points)

Voici une représentation graphique en barres qui correspond, pour la France, au nombre de conducteurs de cyclomoteurs décédés chaque année.

Première partie



Dans toute cette première partie, la lecture graphique sera faîte sans justification.

1. Quelle est l'étendue de cette série statistique ?

2. Est-il vrai qu'entre 2012 et 2022 le nombre de décès à diminué de plus de 45 % ?

3. Calculer la moyenne à l'unité près de cette série statistique.

4. Déterminer une médiane de cette série statistique.

Deuxième partie

Voici la série statistiques du nombre total de décès dans un accident de la circulation en France.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Années	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total
2	Nombre de décès	3 384	3 461	3 477	3 448	3 248	3 244	2 541	2 944	3 267	3 106	32 120

1. Calculer la moyenne de cette série statistique.

2. Quelle formule a été saisie dans la cellule **L2** ?

3. Quelle est l'étendue de cette série statistique ?

4. Déterminer la médiane de cette série.

Exercice n° 3

(6 points)

Lors d'un sondage, on a demandé à des familles le nombre total d'écrans disponibles à la maison en comptant les télévisions, les tablettes et les téléphones.
Voici le résultat :

Nombre d'écrans	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Effectifs	51	134	678	789	1 089	2 678	2 876	1 678	1 089	890	654	145	76	17

1. Combien de familles ont été interrogées pour ce sondage ?

2. Quelle est l'étendue de cette série statistique ?

3. Calculer la moyenne, arrondie à l'unité près, du nombre d'écrans possédés par ces familles.

4. Est-il vrai que la moitié de ces familles possèdent plus de 7 écrans ?

Évaluation — Correction

Exercice n° 1 — Correction

1.

$f(x) = 5x(3x-1) - 7(2x-3) - 8x^2$

$f(x) = 15x^2 - 5x - 14x + 21 - 8x^2$

$\boxed{f(x) = 7x^2 - 19x + 21}$

2.

$g(x) = (3x-7)(2x+3) + (3x-7)(4x-5)$

$g(x) = 6x^2 + 9x - 14x - 21 + 12x^2 - 9x - 28x + 35$

$\boxed{g(x) = 18x^2 - 48x + 14}$

3. Factoriser $g(x)$.

$g(x) = (3x-7)(2x+3) + (3x-7)(4x-5)$

$g(x) = (3x-7)\,[(2x+3) + (4x-5)]$

$g(x) = (3x-7)(2x+3+4x-5)$

$\boxed{g(x) = (3x-7)(6x-2)}$

Exercice n° 2 — Correction

Première partie

1. La valeur minimum est 96 en 2021. La valeur maximum est 179 en 2012.

$\boxed{\text{L'étendue de cette série statistique vaut } 179 - 96 = 83.}$

2. En 2012 il y a environ 179 décès. En 2022 environ 98.
Comme $179 - 98 = 81$, il y a une baisse de 81 décès.

Or $\dfrac{81}{179} \approx 0,4525$ soit $45,3$ %.

$\boxed{\text{Oui, il est vrai que le nombre de décès a baissé de 45 \%.}}$

3. Il faut calculer :

$$\frac{179+159+165+155+121+117+133+134+100+96+98}{11} = \frac{1457}{11} \approx 133$$

$\boxed{\text{En moyenne, il y a eu chaque année environ 123 décès, à l'unité près.}}$

4. Il y a 11 valeurs dans cette série statistique. Comme $11 = 5 + 1 + 5$, la médiane est la sixième valeur dans l'ordre croissant.
Classons les valeurs de la série : $96 < 98 < 100 < 117 < 121 < 133 < 134 < 155 < 159 < 165 < 179$.

$\boxed{\text{La médiane de cette série statistique est 133.}}$

Deuxième partie

Voici la série statistiques du nombre total de décès dans un accident de la circulation en France.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Années	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total
2	Nombre de décès	3 384	3 461	3 477	3 448	3 248	3 244	2 541	2 944	3 267	3 106	32 120

1. Calculons $\dfrac{3384+3461+3477+3448+3248+3244+2541+2944+3267+3104}{10} = \dfrac{32\,120}{10} = 3112$

La moyenne de cette série statistique vaut 3212.

2. Dans la cellule **L2** on a saisi : **=B2+C2+D2+E2+F2+G2+H2+I2+J2+K2** ou **=SOMME(B2:K2)**.

3. La valeur maximale vaut 3477 et la valeur minimale 2541.

L'étendue de cette série vaut : $3477 - 2541 = 936$.

4. Il y a 10 valeurs dans cette série. La médiane est la moyenne de la cinquième et la sixième valeur classée dans l'ordre croissant.
Voici le classement : $2541 < 2944 < 3106 < 3244 < 3248 < 3267 < 3384 < 3448 < 3461 < 3477$

La cinquième valeur vaut 3248 et la sixième 3267. La médiane vaut $\dfrac{3248+3267}{2} = 3257{,}5$

Exercice n° 3 — Correction

1. Calculons $51+134+678+789+1089+2678+2876+1678+1089+890+654+145+76+17 = 12\,844$.

12 844 familles ont été interrogées.

2. L'étendue de cette série vaut $13 - 0 = 13$.

3. Calculons :

$$\frac{51\times0+134\times1+678\times2+789\times3+1089\times4+2678\times5+2876\times6+1678\times7+1089\times8+890\times9+654\times10+145\times11+76\times12+17\times13}{12\,844}$$

$$\frac{76\,595}{12\,844} \approx 5{,}96$$

Ces familles possèdent en moyenne 6 écrans.

4. Il faut déterminer la médiane en déterminant les effectifs cumulés croissants.

L'effectif total vaut 12 844. Comme $12\,844 \div 2 = 6422$, la médiane se situe entre la 6422 ^{e}et 6423 ^{e}valeur.
Calculons les effectifs cumulés croissants.

Nombre d'écrans	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Effectifs	51	134	678	789	1 089	2 678	2 876	1 678	1 089	890	654	145	76	17
.C.C	51	185	863	1 652	2 741	5 419	8 295	...	...	...	...	...	...	...

La valeur cherchée se trouve dans la colonne qui correspond à 6 écrans.

C'est faux ! La moitié des familles possèdent plus de 6 écrans !

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'individus concernés 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, ... x_n est :

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

La **moyenne pondérée** de la série de n valeurs x_1, x_2, x_2, ... x_n pondérées par leurs effectifs respectifs a_1, a_2, a_3, ... a_n est :

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

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 la valeur minimale de cette 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 un nombre 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^{e} et $n+1^{\text{e}}$ valeur.
Tout nombres compris entre la n^{e} et la $n+1^{\text{e}}$ valeur est une médiane dans ce cas.

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 78ᵉ et 79ᵉ valeurs.
D'après le tableau cette médiane se situe dans l'intervalle $[5;10[$.

Informations légales

— **Auteur :** Fabrice ARNAUD
— **Web :** pi.ac3j.fr
— **Mail :** contact@ac3j.fr

— **Dernière modification :** 23 juin 2024 à 16:22

Ce document a été écrit pour LaTeXavec l'éditeur VIM - Vi Improved Vim 9.1.
Il a été compilé sous Linux Ubuntu Noble Numbat 24.04 avec la distribution TeX Live 2023.20240207-101 et LuaHBTex 1.17.0

Pour compiler ce document, un fichier comprennant la plupart des macros est nécessaires. Ce fichier, Entete.tex, est encore trop mal rédigé pour qu'il puisse être mis en ligne. Il est en cours de réécriture et permettra ensuite le partage des sources dans de bonnes conditions.
Le fichier source a été réalisé sous Linux Ubuntu avec l'éditeur Vim. Il utilise une balise spécifique à Vim pour permettre une organisation du fichier sous forme de replis. Cette balise %{{{ ... %}}} est un commentaire pour LaTeX, elle n'est pas nécessaire à sa compilation. Vous pouvez l'utiliser avec Vim en lui précisant que ce code defini un repli. Je vous laisse consulter la documentation officielle de Vim à ce sujet.

Licence CC BY-NC-SA 4.0

Comment créditer cette Œuvre ?

Ce document, **Cours.pdf**, a été crée par **Fabrice ARNAUD (contact@ac3j.fr)** le 23 juin 2024 à 16:22.
Il est disponible en ligne sur **pi.ac3j.fr**, **Le blog de Fabrice ARNAUD**.
Adresse de l'article : **`https://pi.ac3j.fr/mathematiques-college`**.