

Information chiffrée : statistique bivariée

Fiche d'exercices corrigés — 1re ens. sci.

Difficulté : ★ application directe · ★★ standard · ★★★ approfondissement · ★★★★★ défi.

Un conseil : cherchez d'abord, le corrigé ensuite — c'est là que ça progresse.

Exercice 1 ★ Lire et compléter un tableau croisé

Notions : Tableau croisé d'effectifs et fréquences

Un lycée compte 600 élèves. L'intendance étudie le régime des élèves (demi-pensionnaire ou externe) selon leur niveau de classe (seconde, première ou terminale). Le tableau croisé d'effectifs ci-dessous, partiellement rempli, résume la situation.

	Seconde	Première	Terminale	Total
Demi-pensionnaires	150	120	...	380
Externes	...	80	70	...
Total	220	...	180	600

1. Dans cette étude, quelle est la population ? Quels sont les deux caractères observés sur chaque individu ?
2. Que représente le nombre 150 ? Et le nombre 380 ? Répondre par une phrase pour chacun.
3. Recopier et compléter le tableau, en expliquant par un calcul comment on obtient chacune des quatre cases manquantes.
4. Répondre par lecture du tableau complété :
 - a. Combien d'élèves de terminale sont demi-pensionnaires ?
 - b. Combien le lycée compte-t-il d'élèves externes en tout ?
5. Dans quel niveau trouve-t-on le plus d'élèves externes ? Un élève affirme : « C'est normal, c'est le niveau qui compte le plus d'élèves. » A-t-il raison ? Justifier à l'aide du tableau.

Corrigé

1. La population étudiée est l'ensemble des 600 élèves du lycée : chaque élève est un individu de cette population. Sur chaque individu, on observe deux caractères :

- le **régime**, dont les modalités sont « demi-pensionnaire » et « externe » ;
- le **niveau de classe**, dont les modalités sont « seconde », « première » et « terminale ».

Ces deux caractères sont qualitatifs : leurs modalités sont des catégories, et non des nombres.

2. Le nombre 150 est situé à l'intersection de la ligne « Demi-pensionnaires » et de la colonne « Seconde » : c'est un effectif croisé. Il signifie que 150 élèves du lycée sont à la fois en seconde et demi-pensionnaires.

Le nombre 380 est situé dans la colonne « Total », sur la ligne « Demi-pensionnaires » : c'est un effectif marginal. Il signifie que le lycée compte en tout 380 élèves demi-pensionnaires, tous niveaux confondus.

3. Dans un tableau croisé, le total d'une ligne est la somme des cases de cette ligne, et le total d'une colonne est la somme des cases de cette colonne. On en déduit chacune des quatre cases manquantes.

- *Demi-pensionnaires de terminale* : sur la ligne « Demi-pensionnaires », on doit avoir $150 + 120 + \dots = 380$, donc la case vaut $380 - 150 - 120 = 110$. (On peut le vérifier avec la colonne « Terminale » : $180 - 70 = 110$.)
- *Externes de seconde* : dans la colonne « Seconde », on doit avoir $150 + \dots = 220$, donc la case vaut $220 - 150 = 70$.
- *Total de la colonne « Première »* : $120 + 80 = 200$. (Autre méthode, avec la ligne des totaux : $600 - 220 - 180 = 200$.)
- *Total de la ligne « Externes »* : $70 + 80 + 70 = 220$. (Autre méthode : $600 - 380 = 220$.)

D'où le tableau complété :

	Seconde	Première	Terminale	Total
Demi-pensionnaires	150	120	110	380
Externes	70	80	70	220
Total	220	200	180	600

Vérification : $380 + 220 = 600$ et $220 + 200 + 180 = 600$: les marges redonnent bien l'effectif total du lycée.

4. Par lecture du tableau complété :

a. À l'intersection de la ligne « Demi-pensionnaires » et de la colonne « Terminale », on lit 110 : le lycée compte 110 élèves de terminale demi-pensionnaires.

b. Dans la marge de la ligne « Externes », on lit 220 : le lycée compte 220 élèves externes en tout.

5. Sur la ligne « Externes », le plus grand effectif est celui de la colonne « Première » : 80 externes, contre 70 en seconde et 70 en terminale. C'est donc en première que l'on trouve le plus d'élèves externes.

L'élève a tort. D'après la ligne « Total », le niveau qui compte le plus d'élèves est la seconde, avec 220 élèves (contre 200 en première et 180 en terminale). Le plus grand nombre d'externes ne se trouve donc pas dans le niveau le plus nombreux : l'explication proposée par cet élève ne correspond pas à ce que montre le tableau.

Exercice 2 ★★ Fréquences marginales et conditionnelles

Notions : Tableau croisé d'effectifs et fréquences

Dans le cadre d'une enquête de santé publique, on a interrogé 500 adultes d'une même ville. Pour chaque personne, on a relevé deux caractères : pratique-t-elle une activité physique régulière (au moins 30 minutes par jour) ? Se déclare-t-elle en bonne santé ? Les résultats sont regroupés dans le tableau croisé suivant.

	Se déclarent en bonne santé	Ne se déclarent pas en bonne santé	Total
Activité physique régulière	160	40	200
Pas d'activité régulière	180	120	300
Total	340	160	500

1. Calculer la fréquence marginale des personnes qui pratiquent une activité physique régulière, puis celle des personnes qui se déclarent en bonne santé. Exprimer chaque fréquence en pourcentage et l'interpréter par une phrase.
2. Parmi les personnes qui pratiquent une activité physique régulière, quelle est la fréquence de celles qui se déclarent en bonne santé ? Même question parmi les personnes qui ne pratiquent pas d'activité régulière. Comparer ces deux fréquences conditionnelles par une phrase.
3. À partir de cette enquête, un site internet titre : « 80 % des personnes en bonne santé pratiquent une activité physique régulière. » Ce titre est-il correct ? Justifier par un calcul et, si nécessaire, proposer un titre exact.
4. Ces données suffisent-elles à affirmer que l'activité physique régulière est la *cause* de la bonne santé ? Répondre en quelques phrases.

Corrigé

1. Une fréquence marginale s'obtient en divisant un effectif marginal (un total de ligne ou de colonne) par l'effectif total de la population, ici 500 personnes.

Pour l'activité physique régulière, l'effectif marginal est 200 :

$$\frac{200}{500} = 0,40 = 40 \%$$

Interprétation : 40 % des personnes interrogées pratiquent une activité physique régulière.

Pour la bonne santé déclarée, l'effectif marginal est 340 :

$$\frac{340}{500} = 0,68 = 68 \%$$

Interprétation : 68 % des personnes interrogées se déclarent en bonne santé.

2. On calcule cette fois des fréquences conditionnelles : on ne divise plus par 500, mais par l'effectif du groupe auquel on se restreint.

Parmi les 200 personnes qui pratiquent une activité physique régulière, 160 se déclarent en bonne santé :

$$\frac{160}{200} = 0,80 = 80 \%$$

Parmi les 300 personnes qui ne pratiquent pas d'activité régulière, 180 se déclarent en bonne santé :

$$\frac{180}{300} = 0,60 = 60 \%$$

Comparaison : la proportion de personnes se déclarant en bonne santé est nettement plus élevée chez les pratiquants (80 %) que chez les non-pratiquants (60 %) : l'écart est de 20 points de pourcentage.

3. Ce titre n'est pas correct. Il parle « des personnes en bonne santé » : le groupe de référence est donc l'ensemble des 340 personnes qui se déclarent en bonne santé. Parmi elles, 160 pratiquent une activité physique régulière, d'où la fréquence conditionnelle :

$$\frac{160}{340} \approx 0,47$$

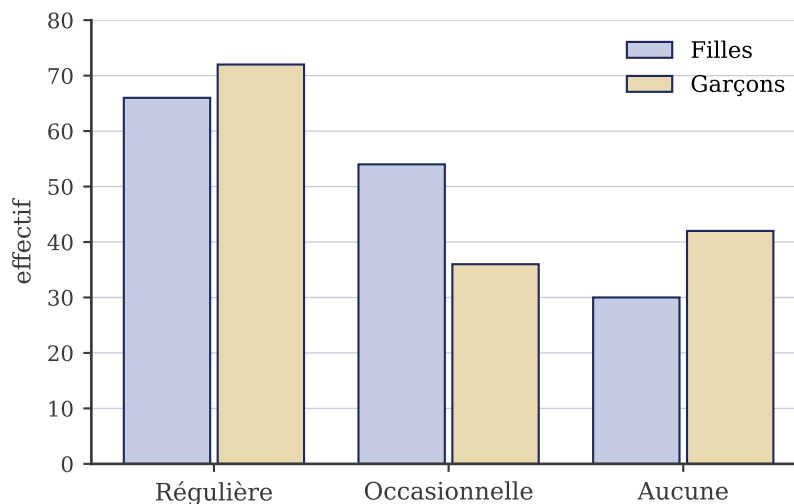
Environ 47 % seulement des personnes en bonne santé pratiquent une activité physique régulière, et non 80 %. Le site a inversé les deux groupes de référence : 80 % est la fréquence des personnes en bonne santé *parmi les pratiquants*, calculée à la question 2., et non la fréquence des pratiquants parmi les personnes en bonne santé. Un titre exact serait : « 80 % des personnes qui pratiquent une activité physique régulière se déclarent en bonne santé », ou encore : « Environ 47 % des personnes en bonne santé pratiquent une activité physique régulière ».

4. Non, ces données ne suffisent pas. L'enquête met en évidence un **lien apparent** entre les deux caractères : la bonne santé déclarée est plus fréquente chez les pratiquants que chez les non-pratiquants. Mais un lien apparent n'est pas une preuve de cause à effet. Le lien pourrait d'abord jouer dans l'autre sens : les personnes déjà en bonne santé ont sans doute plus de facilité à pratiquer une activité physique. Un troisième facteur, comme l'âge ou l'alimentation, peut aussi influencer à la fois la pratique sportive et la santé. Enfin, il s'agit de santé *déclarée* par les personnes elles-mêmes, et non mesurée par un médecin. On ne peut donc pas affirmer, à partir de cette seule enquête, que l'activité physique régulière est la cause de la bonne santé.

Exercice 3 ★ Du diagramme en barres au tableau croisé

Notions : Diagrammes en barres et circulaires, Tableau croisé d'effectifs et fréquences

Une enquête sur la pratique sportive a été menée auprès des 300 élèves de seconde d'un lycée : 150 filles et 150 garçons. Chaque élève a déclaré une pratique sportive « régulière », « occasionnelle » ou « aucune ». Les résultats sont représentés par le diagramme en barres groupées ci-dessous.



On sait que les six effectifs représentés sont, dans le désordre : 30, 36, 42, 54, 66 et 72.

- À l'aide du quadrillage, associer chaque barre du diagramme à son effectif. On expliquera par exemple comment distinguer les deux barres de la catégorie « Régulière ».
- Reconstruire le tableau croisé d'effectifs de l'enquête, avec sa ligne et sa colonne « Total ». Vérifier que l'on retrouve bien 150 filles, 150 garçons et 300 élèves.
- Calculer, en pourcentage :
 - la fréquence marginale des élèves sans aucune pratique sportive ;
 - parmi les filles, la fréquence de la pratique régulière ;
 - parmi les garçons, la fréquence de la pratique régulière.
- Un journaliste du lycée conclut : « Les garçons sont plus sportifs que les filles. » À l'aide du diagramme ou du tableau, expliquer pourquoi cette conclusion mérite d'être nuancée.

Corrigé

1. Le quadrillage comporte une ligne horizontale tous les 10 élèves : on lit chaque effectif en repérant la position du sommet de la barre par rapport à ces lignes.

Pour la catégorie « Régulière », les deux barres arrivent au voisinage de la ligne des 70 : celle des filles s'arrête juste en dessous, tandis que celle des garçons la dépasse légèrement. Or les deux seuls effectifs de la liste proches de 70 sont 66 et 72 : la barre des filles correspond donc à 66 et celle des garçons à 72.

On procède de même pour les deux autres catégories :

- « Occasionnelle » : la barre des filles s'arrête entre les lignes 50 et 60, c'est donc 54 ; celle des garçons entre les lignes 30 et 40, c'est donc 36 ;
- « Aucune » : la barre des filles s'arrête exactement sur la ligne des 30, c'est donc 30 ; celle des garçons un peu au-dessus de la ligne des 40, c'est donc 42.

2. On reporte ces six effectifs dans un tableau croisé, puis on calcule les marges en additionnant

les cases de chaque ligne et de chaque colonne :

	Régulière	Occasionnelle	Aucune	Total
Filles	66	54	30	150
Garçons	72	36	42	150
Total	138	90	72	300

Vérification : $66 + 54 + 30 = 150$ et $72 + 36 + 42 = 150$: on retrouve bien les 150 filles et les 150 garçons annoncés, soit $150 + 150 = 300$ élèves en tout. (On peut aussi contrôler la dernière ligne : $138 + 90 + 72 = 300$.)

3. On divise chaque effectif par celui du groupe de référence : les 300 élèves pour la fréquence marginale, les 150 filles ou les 150 garçons pour les fréquences conditionnelles.

a. $\frac{72}{300} = 0,24$, soit 24 % : environ un quart des élèves de seconde n'a aucune pratique sportive.

b. $\frac{66}{150} = 0,44$, soit 44 % : parmi les filles, 44 % déclarent une pratique régulière.

c. $\frac{72}{150} = 0,48$, soit 48 % : parmi les garçons, 48 % déclarent une pratique régulière.

4. La conclusion mérite d'être nuancée, car elle dépend de la colonne du tableau que l'on regarde.

— Pour la pratique **régulière**, les garçons devancent effectivement les filles : 48 % contre 44 % (question **3.**).

— Mais les garçons sont aussi les plus nombreux à n'avoir **aucune** pratique : $\frac{42}{150} = 0,28$, soit 28 % des garçons, contre $\frac{30}{150} = 0,20$, soit 20 % des filles.

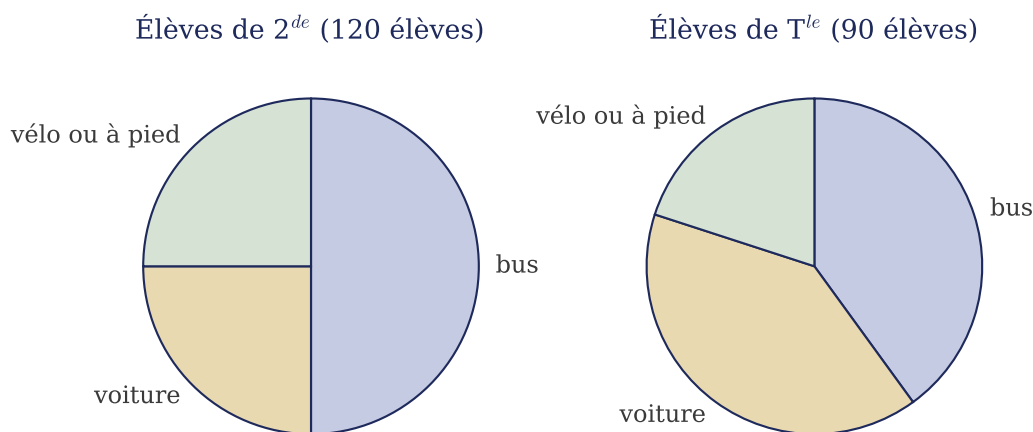
— Enfin, si l'on regroupe pratiques régulière et occasionnelle : $\frac{66 + 54}{150} = \frac{120}{150} = 0,80$, soit 80 % des filles pratiquent un sport, contre $\frac{72 + 36}{150} = \frac{108}{150} = 0,72$, soit 72 % des garçons.

Selon le critère choisi, ce sont donc tantôt les garçons (pratique régulière), tantôt les filles (pratique sportive en général) qui apparaissent « plus sportifs ». La conclusion du journaliste est trop rapide : elle ne s'appuie que sur une partie des données.

Exercice 4 ★★ Comparer deux diagrammes circulaires

Notions : Diagrammes en barres et circulaires

On a demandé aux élèves de seconde et de terminale d'un lycée leur principal moyen de transport pour venir en cours : « bus », « voiture » ou « vélo ou à pied ». Les réponses des 120 élèves de seconde et des 90 élèves de terminale sont représentées par les deux diagrammes circulaires ci-dessous, sur lesquels les effectifs ne sont pas indiqués.



- Sur le diagramme des élèves de seconde, le secteur « bus » occupe la moitié du disque, et les secteurs « voiture » et « vélo ou à pied » un quart chacun. En déduire l'effectif correspondant à chaque moyen de transport chez les élèves de seconde.
- Pour les élèves de terminale, on dispose du tableau d'effectifs suivant :

Moyen de transport	Bus	Voiture	Vélo ou à pied
Effectif	36	36	18

- Calculer la proportion, en pourcentage, de chaque moyen de transport chez les élèves de terminale.
 - Expliquer pourquoi ces proportions sont cohérentes avec le diagramme des terminales.
- Calculer l'angle du secteur « vélo ou à pied » du diagramme des terminales, puis celui du secteur « bus » du diagramme des secondes.
 - Un élève affirme : « Le secteur «voiture» des terminales est bien plus grand que celui des secondes : il y a donc beaucoup plus d'élèves de terminale que de seconde qui viennent en voiture. » Que penser de cette affirmation ? Justifier en comparant à la fois les proportions et les effectifs.
 - Dans quel niveau la proportion d'élèves venant à vélo ou à pied est-elle la plus élevée ? Écrire une phrase de comparaison qui précise bien les deux proportions.

Corrigé

- Dans un diagramme circulaire, la part du disque occupée par un secteur est égale à la proportion de la modalité dans le groupe. On applique donc les fractions données à l'effectif total des secondes, soit 120 élèves.

- Bus : la moitié du disque, soit $\frac{120}{2} = 60$ élèves ;
- Voiture : le quart du disque, soit $\frac{120}{4} = 30$ élèves ;
- Vélo ou à pied : le quart du disque également, soit 30 élèves.

Vérification : $60 + 30 + 30 = 120$: on retrouve bien les 120 élèves de seconde.

2. a. L'effectif total des terminales est $36 + 36 + 18 = 90$ élèves. On divise chaque effectif par 90 :

- bus : $\frac{36}{90} = 0,40$, soit 40 % des élèves de terminale ;
- voiture : $\frac{36}{90} = 0,40$, soit 40 % également ;
- vélo ou à pied : $\frac{18}{90} = 0,20$, soit 20 %.

Vérification : $40 + 40 + 20 = 100$: les trois proportions représentent bien la totalité du groupe.

b. Sur le diagramme des terminales, les secteurs « bus » et « voiture » sont identiques : c'est cohérent, puisque ces deux moyens de transport ont la même proportion, 40 %, soit deux cinquièmes du disque chacun. Le secteur « vélo ou à pied » est deux fois plus petit que chacun des deux autres : c'est cohérent avec sa proportion de 20 %, un cinquième du disque. Les proportions calculées correspondent donc bien au diagramme.

3. Dans un diagramme circulaire, l'angle d'un secteur est proportionnel à la fréquence de la modalité : le disque entier, soit 360° , représente 100 % du groupe.

Secteur « vélo ou à pied » des terminales : la fréquence est 0,20, donc l'angle vaut $0,20 \times 360^\circ = 72^\circ$.

Secteur « bus » des secondes : la fréquence est 0,50 (la moitié du groupe), donc l'angle vaut $0,50 \times 360^\circ = 180^\circ$. C'est un demi-disque, ce qui est bien cohérent avec la figure.

4. L'affirmation est fausse, car elle confond proportions et effectifs.

- En **proportion**, la voiture pèse effectivement davantage chez les terminales : 40 %, contre $\frac{30}{120} = 0,25$, soit 25 % chez les secondes. C'est ce que montrent les tailles des secteurs.
- En **effectif**, 36 élèves de terminale viennent en voiture, contre 30 élèves de seconde : l'écart n'est que de $36 - 30 = 6$ élèves, ce qui n'est pas « beaucoup plus ».

Un diagramme circulaire représente des proportions à l'intérieur d'un groupe. Comme les deux groupes n'ont pas le même effectif total (90 terminales contre 120 secondes), comparer directement les tailles des secteurs de deux diagrammes ne renseigne pas sur les effectifs : un secteur plus grand dans un groupe plus petit peut représenter à peine plus d'individus.

5. Chez les secondes, la proportion d'élèves venant à vélo ou à pied est $\frac{30}{120} = 0,25$, soit 25 % ; chez les terminales, elle est de 20 % (question **2. a.**). *Phrase de comparaison* : la proportion d'élèves venant à vélo ou à pied est plus élevée chez les élèves de seconde, où elle atteint 25 %, que chez les élèves de terminale, où elle vaut 20 %.

Exercice 5 ★ Construire un nuage de points

Notions : Nuage de points

Dans le cadre d'une étude sur le sommeil, huit adolescents ont noté leur temps d'écran quotidien x (en heures par jour) et leur durée moyenne de sommeil y (en heures par nuit). Voici les résultats :

Temps d'écran x (h/jour)	1	2	2	3	4	5	6	7
Sommeil y (h/nuit)	9,5	9	8,5	8,5	8	7,5	7	6,5

1. On veut représenter cette série par un nuage de points. Proposer un repère adapté : quelle grandeur placer en abscisse, quelle grandeur en ordonnée, et quelles graduations choisir sur chaque axe ? On expliquera pourquoi il est inutile de faire commencer l'axe des ordonnées à 0.
2. Construire le nuage de points sur papier quadrillé. Combien de points le nuage compte-t-il ?
3. Deux adolescents de l'étude passent exactement 2 heures par jour devant un écran. Comment cela se traduit-il sur le nuage ?
4. Décrire la tendance du nuage : que semble-t-il se passer pour la durée de sommeil lorsque le temps d'écran augmente ?
5. Une élève conclut : « Ce graphique prouve que les écrans font perdre du sommeil. » Expliquer en quelques phrases pourquoi cette conclusion est trop rapide.

Corrigé

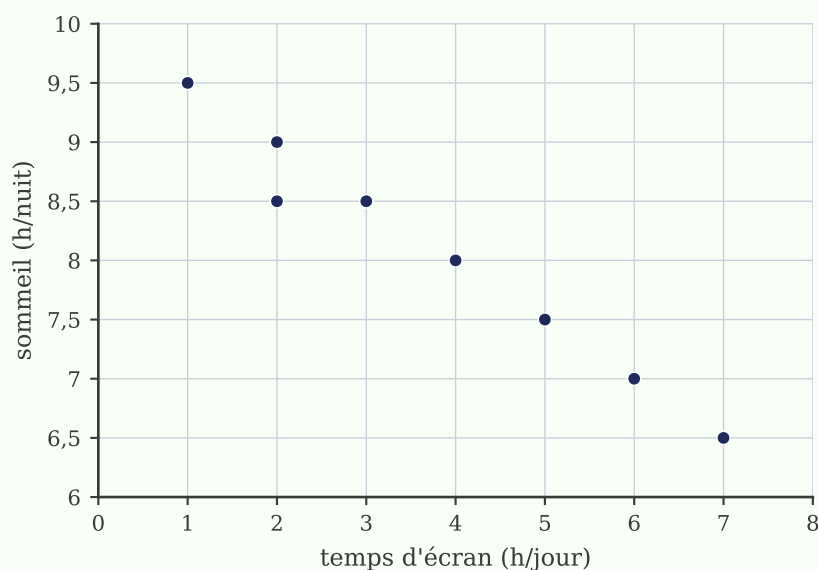
1. On étudie la durée de sommeil *en fonction du* temps d'écran : on place donc le temps d'écran x en abscisse et la durée de sommeil y en ordonnée.

Pour les graduations : les temps d'écran vont de 1 à 7 h, on peut donc graduer l'axe des abscisses de 0 à 8, d'heure en heure. Les durées de sommeil vont de 6,5 à 9,5 h : on peut graduer l'axe des ordonnées de 6 à 10, de 0,5 en 0,5.

Il est inutile de faire commencer l'axe des ordonnées à 0 : toutes les valeurs de y sont comprises entre 6,5 et 9,5 h. Si l'axe partait de 0, toute la zone comprise entre 0 et 6 h resterait vide, et les huit points seraient écrasés dans le haut du graphique : le nuage deviendrait illisible.

2. À chaque adolescent correspond un point dont l'abscisse est son temps d'écran et l'ordonnée sa durée de sommeil. Comme l'étude porte sur huit adolescents, le nuage compte 8 points :

$$(1; 9,5), (2; 9), (2; 8,5), (3; 8,5), (4; 8), (5; 7,5), (6; 7), (7; 6,5)$$



3. Ces deux adolescents ont le même temps d'écran, $x = 2$, mais des durées de sommeil différentes : 9 h pour l'un, 8,5 h pour l'autre. Ils donnent donc deux points de même abscisse, (2 ; 9) et (2 ; 8,5), situés l'un au-dessus de l'autre sur la même verticale. Cela montre qu'à temps d'écran égal, la durée de sommeil peut différer d'une personne à l'autre.

4. Quand on parcourt le nuage de gauche à droite, les points descendent régulièrement en restant presque alignés : la tendance du nuage est décroissante. Autrement dit, dans cette étude, plus le temps d'écran quotidien d'un adolescent est grand, plus sa durée moyenne de sommeil est courte.

5. Cette conclusion est trop rapide. Le nuage montre un **lien apparent** entre le temps d'écran et la durée de sommeil, mais un lien apparent n'est pas une preuve de cause à effet. Le lien pourrait d'abord jouer dans l'autre sens : ce sont peut-être les adolescents qui dorment peu qui occupent leurs soirées devant un écran. Une troisième grandeur, comme le rythme de vie, l'âge ou le stress, peut aussi agir à la fois sur le temps d'écran et sur le sommeil. Enfin, l'étude ne porte que sur huit adolescents : un échantillon aussi petit ne permet pas de « prouver » quoi que ce soit sur l'ensemble des adolescents.

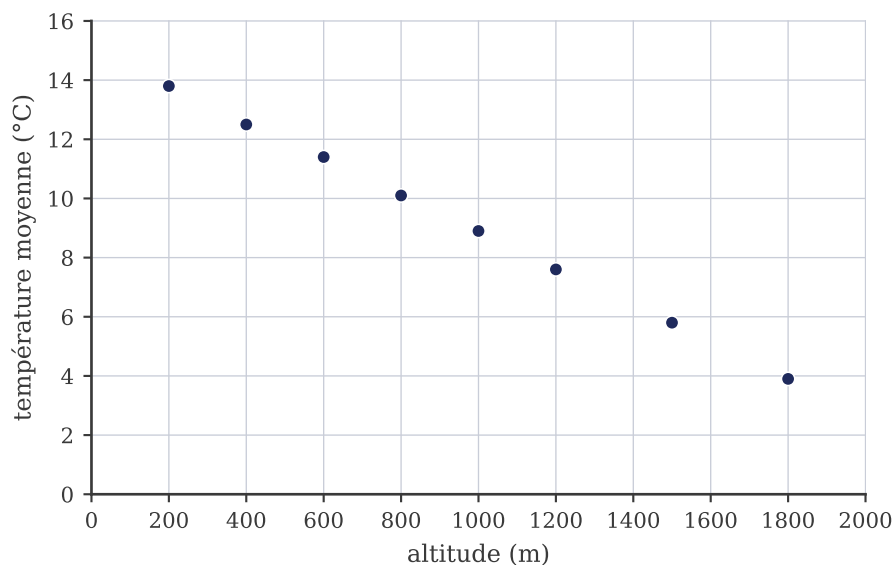
Exercice 6 ★ Décrire un nuage de points

Notions : Nuage de points

Huit stations météo d'un même massif montagneux ont relevé leur température moyenne annuelle. Le tableau ci-dessous donne, pour chaque station, son altitude x (en mètres) et sa température moyenne annuelle y (en °C).

Altitude x (m)	200	400	600	800	1 000	1 200	1 500	1 800
Température y (°C)	13,8	12,5	11,4	10,1	8,9	7,6	5,8	3,9

Le nuage de points correspondant est représenté ci-dessous.



- Dans cette série, les deux caractères observés sont-ils qualitatifs ou quantitatifs ? Expliquer pourquoi un nuage de points est ici mieux adapté qu'un tableau croisé d'effectifs.
- Répondre par lecture graphique :
 - Quelle est la température moyenne annuelle de la station située à 800 m d'altitude ?
 - À quelle altitude se trouve la station dont la température moyenne annuelle est 5,8 °C ?
- Décrire la tendance du nuage par une phrase : que se passe-t-il pour la température lorsque l'altitude augmente ?
- De combien la température moyenne baisse-t-elle, environ, quand on monte de 400 m ? On appuiera la réponse sur au moins deux exemples pris dans le tableau.
- Un bulletin météo affirme : « En montagne, on perd environ 6 °C chaque fois que l'on s'élève de 1 000 m. » Ces données confirment-elles cet ordre de grandeur ? Justifier par un calcul.

Corrigé

1. L'altitude et la température moyenne annuelle sont deux caractères **quantitatifs** : leurs valeurs sont des nombres issus d'une mesure.

Un tableau croisé d'effectifs sert à croiser deux caractères qualitatifs : on compte les individus dans chaque case. Ici, les huit stations ont toutes des valeurs différentes : chaque case du tableau contiendrait au plus un individu, et on ne verrait rien. Pour deux caractères quantitatifs, on représente plutôt chaque individu — ici chaque station — par un point de coordonnées $(x; y)$:

le nuage de points ainsi obtenu permet de visualiser d'un coup d'œil la tendance de la série.

2. Par lecture graphique :

- a.** Le point d'abscisse 800 a pour ordonnée 10,1 : la station située à 800 m d'altitude a une température moyenne annuelle de 10,1 °C.
- b.** Le point d'ordonnée 5,8 a pour abscisse 1 500 : la station dont la température moyenne annuelle est 5,8 °C se trouve à 1 500 m d'altitude.

3. Quand on parcourt le nuage de gauche à droite, les points descendent en restant presque alignés : la tendance du nuage est décroissante. Autrement dit, plus l'altitude d'une station est élevée, plus sa température moyenne annuelle est basse.

4. On compare des stations dont les altitudes diffèrent de 400 m :

- de 200 m à 600 m : la température passe de 13,8 °C à 11,4 °C, soit une baisse de $13,8 - 11,4 = 2,4$ °C ;
- de 600 m à 1 000 m : la température passe de 11,4 °C à 8,9 °C, soit une baisse de $11,4 - 8,9 = 2,5$ °C ;
- de 800 m à 1 200 m : la température passe de 10,1 °C à 7,6 °C, soit une baisse de $10,1 - 7,6 = 2,5$ °C.

À chaque fois, la baisse est de 2,4 à 2,5 °C : quand on monte de 400 m, la température moyenne annuelle baisse donc d'environ 2,5 °C.

5. On choisit dans le tableau des stations séparées de 1 000 m de dénivelé :

- de 200 m à 1 200 m d'altitude : $13,8 - 7,6 = 6,2$ °C de baisse ;
- de 800 m à 1 800 m d'altitude : $10,1 - 3,9 = 6,2$ °C de baisse.

Dans les deux cas, la température perd environ 6 °C pour 1 000 m d'ascension : les relevés de ce massif confirment bien l'ordre de grandeur annoncé par le bulletin météo.

Exercice 7 ★ Calculer les coordonnées du point moyen

Notions : Point moyen

Le propriétaire d'un panneau solaire a relevé, pendant six journées, la durée d'ensoleillement x (en heures) et l'énergie électrique produite par le panneau y (en kWh).

Ensoleillement x (h)	2	4	5	7	8	10
Production y (kWh)	4	7	9	12	13	15

1. Calculer la moyenne $\bar{x}$ des durées d'ensoleillement, en détaillant le calcul.
2. Calculer la moyenne $\bar{y}$ des productions, en détaillant le calcul.
3. En déduire les coordonnées du point moyen G du nuage de points associé à cette série.
4. Le point G correspond-il à l'une des six journées relevées ? Justifier.
5. Interpréter les coordonnées de G par une phrase dans le contexte de l'exercice.

Corrigé

1. La moyenne $\bar{x}$ des durées d'ensoleillement s'obtient en additionnant les six valeurs de x , puis en divisant la somme par 6 :

$$\bar{x} = \frac{2 + 4 + 5 + 7 + 8 + 10}{6} = \frac{36}{6} = 6$$

Sur ces six journées, la durée moyenne d'ensoleillement est donc de 6 h par jour.

2. On procède de la même façon avec les six valeurs de y :

$$\bar{y} = \frac{4 + 7 + 9 + 12 + 13 + 15}{6} = \frac{60}{6} = 10$$

Sur ces six journées, la production moyenne du panneau est donc de 10 kWh par jour.

3. Par définition, le point moyen d'un nuage est le point $G(\bar{x}; \bar{y})$, dont l'abscisse est la moyenne des abscisses et l'ordonnée la moyenne des ordonnées. D'après les questions 1. et 2., le point moyen de ce nuage est donc $G(6; 10)$.

4. Non. Pour que G corresponde à l'une des journées relevées, il faudrait qu'une colonne du tableau soit exactement (6; 10). Or aucune des six journées n'a eu une durée d'ensoleillement de 6 h : les valeurs relevées sont 2, 4, 5, 7, 8 et 10 h. Le point G ne correspond donc à aucune journée réelle : le point moyen résume le nuage, mais il n'est pas nécessairement l'un de ses points.

5. Les coordonnées de G s'interprètent ainsi : au cours de ces six journées, le panneau solaire a reçu en moyenne 6 h d'ensoleillement par jour et a produit en moyenne 10 kWh par jour. Le point G représente en quelque sorte la « journée moyenne » de la série, même si, comme on l'a vu à la question 4., aucune journée réelle ne lui correspond exactement.

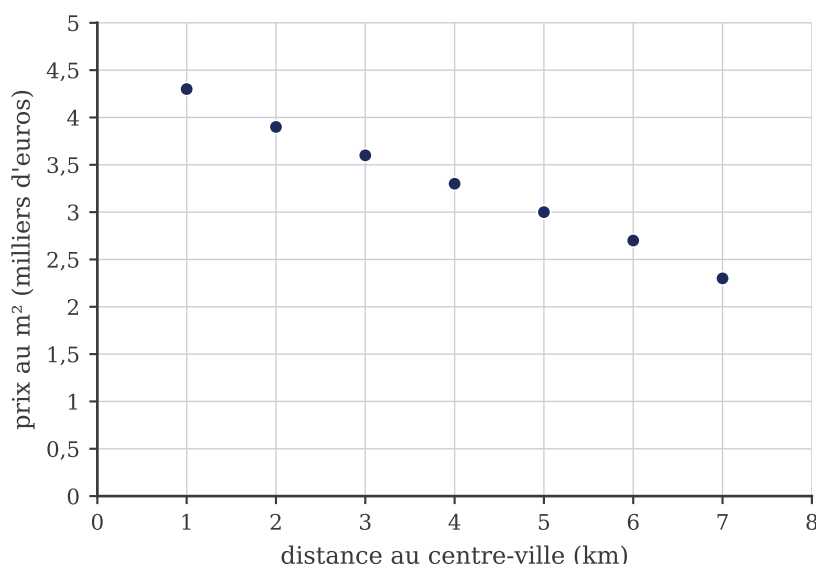
Exercice 8 ★★ Le point moyen dans le nuage

Notions : Nuage de points, Point moyen

Une agence immobilière étudie sept appartements comparables d'une même ville. Pour chacun, elle relève la distance au centre-ville x (en km) et le prix de vente au m^2 y (en milliers d'euros).

Distance x (km)	1	2	3	4	5	6	7
Prix au m^2 y (milliers d'euros)	4,3	3,9	3,6	3,3	3,0	2,7	2,3

Le nuage de points correspondant est représenté ci-dessous.



1. Décrire la tendance du nuage, puis l'interpréter par une phrase dans le contexte de l'exercice.
2. Calculer les coordonnées du point moyen G du nuage, en détaillant les deux calculs de moyenne.
3. Reproduire le graphique et y placer le point G .
4. Que remarque-t-on sur la position de G par rapport aux points du nuage ? Cette situation se produit-elle pour tous les nuages de points ? Justifier, par exemple à l'aide d'un nuage de deux points de son choix.
5. **a.** Exprimer en euros le prix moyen au m^2 des sept appartements, puis interpréter les coordonnées de G par une phrase.
b. Une famille dispose d'un budget maximal de 3 000 € par m^2 . D'après cette série, quels appartements peut-elle envisager ?

Corrigé

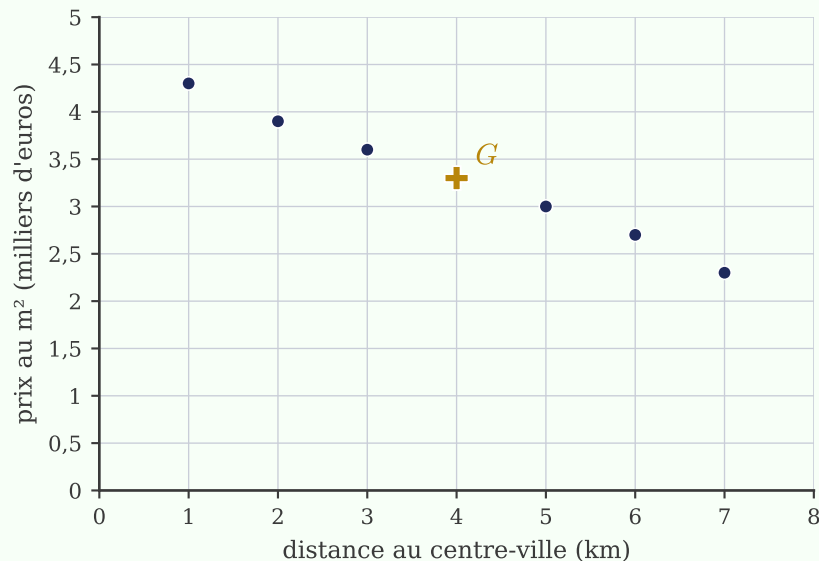
1. Quand on parcourt le nuage de gauche à droite, les points descendent en restant presque alignés : la tendance du nuage est décroissante. *Interprétation* : plus un appartement est éloigné du centre-ville, plus son prix de vente au m^2 est faible.
2. L'abscisse de G est la moyenne $\bar{x}$ des distances, et son ordonnée est la moyenne $\bar{y}$ des prix au m^2 .

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7}{7} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{4,3 + 3,9 + 3,6 + 3,3 + 3,0 + 2,7 + 2,3}{7} = \frac{23,1}{7} = 3,3$$

Le point moyen du nuage est donc $G(\bar{x}; \bar{y})$, soit $G(4; 3,3)$.

3. On place le point $G(4; 3,3)$ sur le graphique : il se situe au « centre d'équilibre » du nuage.



4. On remarque que G est confondu avec l'un des points du nuage : l'appartement situé à 4 km du centre-ville se vend justement 3,3 milliers d'euros le m². C'est une coïncidence, pas une règle générale : le point moyen n'appartient pas toujours au nuage. Par exemple, pour le nuage formé des deux points (0; 1) et (2; 2), le point moyen a pour coordonnées $\bar{x} = \frac{0+2}{2} = 1$ et $\bar{y} = \frac{1+2}{2} = 1,5$: le point $G(1; 1,5)$, milieu du segment joignant les deux points, ne coïncide avec aucun d'eux.

5. a. Le prix moyen au m² est $\bar{y} = 3,3$ milliers d'euros, c'est-à-dire $3,3 \times 1\,000 = 3\,300$ € par m². *Interprétation* : en moyenne, les sept appartements étudiés sont situés à 4 km du centre-ville et se vendent 3 300 € le m².

b. Le budget de la famille impose un prix au m² inférieur ou égal à 3 000 €, c'est-à-dire à 3,0 milliers d'euros. D'après le tableau, trois appartements conviennent : ceux situés à 5 km (3 000 € le m²), à 6 km (2 700 €) et à 7 km (2 300 €) du centre-ville. La famille peut donc envisager les trois appartements les plus éloignés du centre.

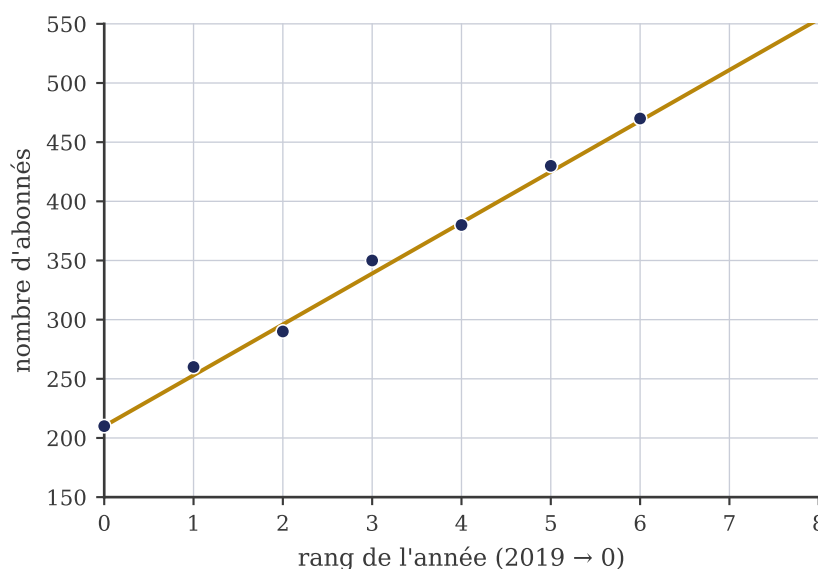
Exercice 9 ★★ Un ajustement au jugé

Notions : Ajustement affine (au jugé, Mayer, moindres carrés), Interpolation et extrapolation

Une salle de sport a relevé son nombre d'abonnés chaque année, de 2019 à 2025. On note x le rang de l'année (l'année 2019 correspond à $x = 0$) et y le nombre d'abonnés.

Année	2019	2020	2021	2022	2023	2024	2025
Rang x	0	1	2	3	4	5	6
Nombre d'abonnés y	210	260	290	350	380	430	470

On a représenté le nuage de points ci-dessous, puis tracé **au jugé** (à la règle, « à l'œil ») une droite d'ajustement $\mathcal{D}$. Cette droite passe par les points $A(0; 210)$ et $B(5; 425)$.



1. Vérifier sur le graphique que la droite $\mathcal{D}$ passe près de tous les points du nuage. Pourquoi peut-on dire qu'un ajustement affine est pertinent ici ?
2. À l'aide des points A et B, calculer le coefficient directeur de la droite $\mathcal{D}$, puis donner son ordonnée à l'origine. En déduire l'équation réduite de $\mathcal{D}$.
3. Interpréter le coefficient directeur dans le contexte de l'exercice.
4. En utilisant l'équation de $\mathcal{D}$, estimer le nombre d'abonnés de la salle en 2027.
5. Cette estimation est-elle une **interpolation** ou une **extrapolation** ? Est-elle fiable ? Justifier.

Corrigé

1. Sur le graphique, la droite $\mathcal{D}$ passe très près de chacun des sept points du nuage : certains points sont légèrement au-dessus de la droite, d'autres légèrement en dessous, mais les écarts restent faibles (quelques abonnés seulement, jamais plus d'une quinzaine). Le nuage de points a donc une forme allongée, presque rectiligne : les points semblent s'aligner le long d'une même droite. C'est exactement dans cette situation qu'un ajustement affine est pertinent : remplacer le nuage par une droite permet de résumer la tendance du phénomène sans trahir les données observées.

2. La droite $\mathcal{D}$ passe par les points $A(0; 210)$ et $B(5; 425)$. Son équation réduite est de la forme $y = mx + p$.

Calcul du coefficient directeur :

$$m = \frac{y_B - y_A}{x_B - x_A} = \frac{425 - 210}{5 - 0} = \frac{215}{5} = 43$$

Calcul de l'ordonnée à l'origine : le point $A(0; 210)$ a pour abscisse 0, c'est donc le point où la droite $\mathcal{D}$ coupe l'axe des ordonnées. On lit directement l'ordonnée à l'origine : $p = 210$.

La droite $\mathcal{D}$ a donc pour équation réduite :

$$y = 43x + 210$$

3. Le coefficient directeur vaut 43 : lorsque x augmente de 1, y augmente de 43. Dans le contexte de l'exercice, cela signifie que, selon ce modèle, **la salle de sport gagne environ 43 abonnés supplémentaires chaque année.**

4. L'année 2019 correspond à $x = 0$, donc l'année 2027 correspond au rang $x = 2027 - 2019 = 8$. On remplace x par 8 dans l'équation de la droite :

$$y = 43 \times 8 + 210 = 344 + 210 = 554$$

Selon ce modèle, la salle compterait **environ 554 abonnés en 2027.**

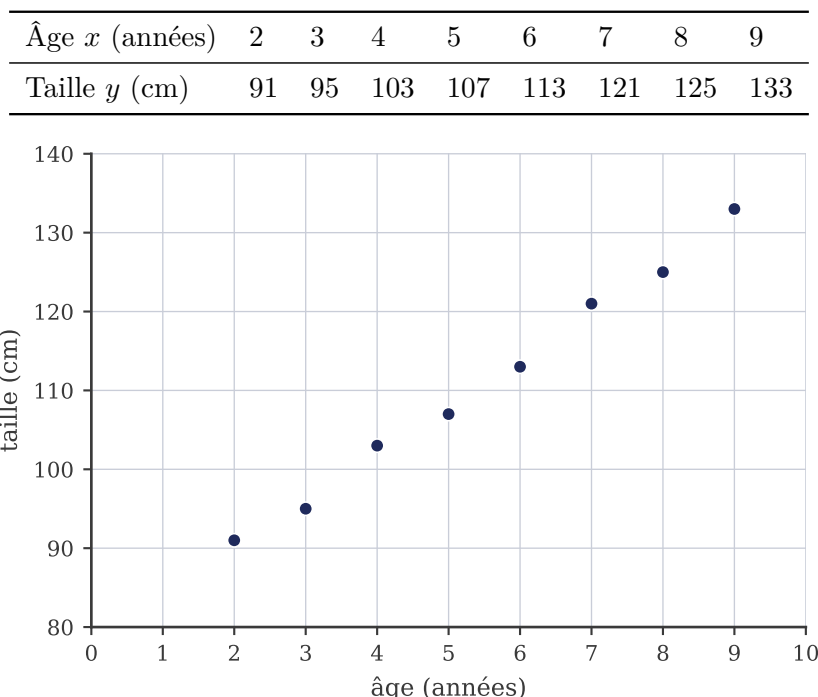
5. Les données observées vont de $x = 0$ (année 2019) à $x = 6$ (année 2025). Or l'estimation de la question 4 utilise $x = 8$, qui se trouve **en dehors** de cette plage : il s'agit donc d'une **extrapolation**.

Sa fiabilité doit être discutée. D'un côté, l'estimation reste proche des données (seulement 2 ans après le dernier relevé) et la tendance observée est très régulière : elle est donc plausible à court terme. De l'autre, elle repose sur l'hypothèse que la croissance se poursuit exactement au même rythme, ce que rien ne garantit : la salle peut atteindre sa capacité maximale, un concurrent peut ouvrir à proximité, les habitudes des clients peuvent changer. Cette estimation donne donc un ordre de grandeur raisonnable, mais elle doit être utilisée avec prudence.

Exercice 10 ★★★ La droite de Mayer

Notions : Point moyen, Ajustement affine (au jugé, Mayer, moindres carrés)

On a relevé la taille d'un enfant à chaque anniversaire, de ses 2 ans à ses 9 ans. On note x l'âge (en années) et y la taille (en cm).



Pour ajuster ce nuage par une droite, on utilise la **méthode de Mayer** : on partage le nuage, rangé dans l'ordre des x croissants, en deux moitiés, on calcule le point moyen de chaque moitié, puis on trace la droite passant par ces deux points.

1. On partage le nuage en deux moitiés : les 4 premiers points d'un côté, les 4 derniers de l'autre. Calculer les coordonnées du point moyen G_1 des 4 premiers points, puis celles du point moyen G_2 des 4 derniers points.
2. Déterminer l'équation réduite de la droite (G_1G_2) , appelée **droite de Mayer** du nuage.
3. Interpréter le coefficient directeur de cette droite dans le contexte de l'exercice.
4. À l'aide de ce modèle, estimer la taille de l'enfant à 10 ans.
5. Que prévoit ce modèle pour la taille de cette personne à 30 ans ? Expliquer pourquoi ce résultat est absurde et quel enseignement on peut en tirer sur l'utilisation d'un ajustement affine.

Corrigé

1. On range les huit points dans l'ordre des âges croissants, puis on partage le nuage en deux moitiés de quatre points chacune.

Point moyen G_1 de la première moitié (âges 2, 3, 4 et 5 ans) : son abscisse est la moyenne des âges et son ordonnée est la moyenne des tailles correspondantes.

$$x_{G_1} = \frac{2 + 3 + 4 + 5}{4} = \frac{14}{4} = 3,5 \quad \text{et} \quad y_{G_1} = \frac{91 + 95 + 103 + 107}{4} = \frac{396}{4} = 99$$

Donc $G_1(3,5; 99)$.

Point moyen G_2 de la seconde moitié (âges 6, 7, 8 et 9 ans) :

$$x_{G_2} = \frac{6 + 7 + 8 + 9}{4} = \frac{30}{4} = 7,5 \quad \text{et} \quad y_{G_2} = \frac{113 + 121 + 125 + 133}{4} = \frac{492}{4} = 123$$

Donc $G_2(7,5; 123)$.

2. La droite de Mayer est la droite (G_1G_2) . Son équation réduite est de la forme $y = mx + p$.
Calcul du coefficient directeur :

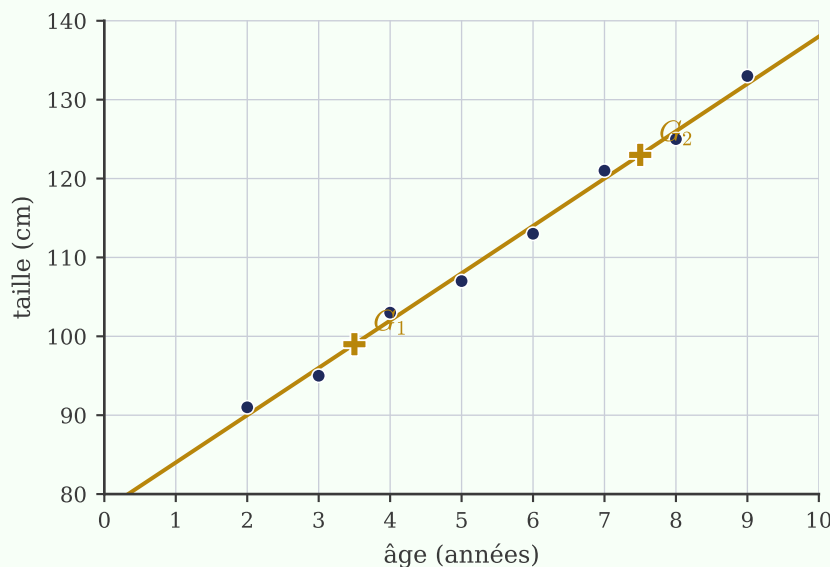
$$m = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{123 - 99}{7,5 - 3,5} = \frac{24}{4} = 6$$

Calcul de l'ordonnée à l'origine : la droite passe par $G_1(3,5; 99)$, donc les coordonnées de G_1 vérifient l'équation $y = 6x + p$:

$$99 = 6 \times 3,5 + p \Leftrightarrow 99 = 21 + p \Leftrightarrow p = 99 - 21 = 78$$

La droite de Mayer a donc pour équation réduite $y = 6x + 78$.

Vérification avec G_2 : $6 \times 7,5 + 78 = 45 + 78 = 123 = y_{G_2}$, donc la droite passe bien aussi par G_2 .



3. Le coefficient directeur vaut 6 : lorsque l'âge augmente de 1 an, la taille donnée par le modèle augmente de 6 cm. Autrement dit, selon ce modèle, **l'enfant grandit d'environ 6 cm par an** entre ses 2 ans et ses 9 ans.

4. Pour estimer la taille de l'enfant à 10 ans, on remplace x par 10 dans l'équation de la droite de Mayer :

$$y = 6 \times 10 + 78 = 60 + 78 = 138$$

On estime donc la taille de l'enfant à 10 ans à **environ 138 cm**. Les données observées vont de 2 à 9 ans : cette estimation est une **extrapolation**, mais elle est très proche de la plage des données (1 an seulement après la dernière mesure), donc elle reste raisonnable.

5. Pour $x = 30$, le modèle donne :

$$y = 6 \times 30 + 78 = 180 + 78 = 258$$

soit une taille de 258 cm, c'est-à-dire 2,58 m ! Ce résultat est absurde : la croissance humaine ralentit puis **s'arrête** à la fin de l'adolescence ; personne ne continue de grandir de 6 cm par an à l'âge adulte. Le phénomène n'est donc plus du tout affine au-delà de l'enfance.

L'enseignement à en tirer est le suivant : un ajustement affine ne décrit correctement le phénomène que **sur la plage des données observées** (ici de 2 à 9 ans) et, éventuellement, tout près de cette plage. C'est une **extrapolation** très éloignée des données qui a produit ce résultat absurde : plus une extrapolation s'éloigne de la plage des données, moins elle est fiable, au point de perdre tout sens.

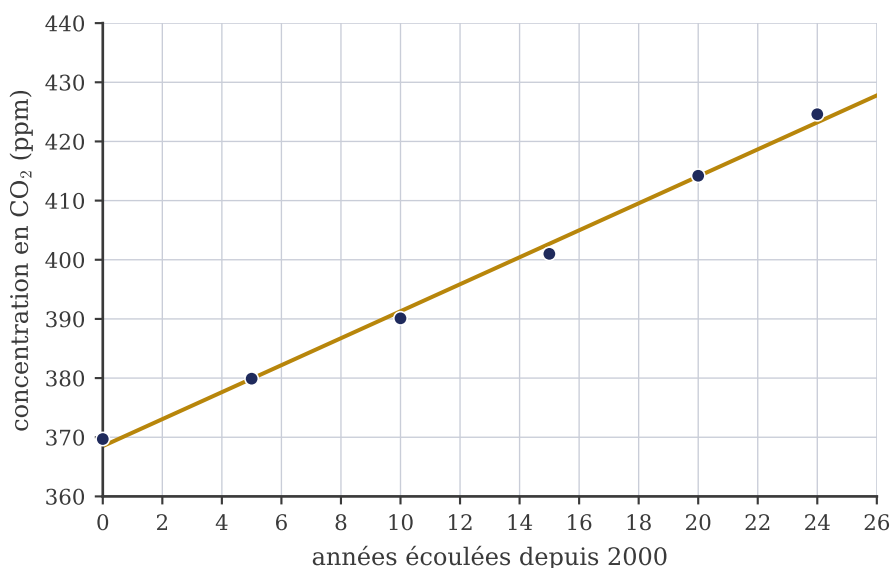
Exercice 11 ★★ La droite des moindres carrés au tableur

Notions : Ajustement affine (au jugé, Mayer, moindres carrés), Interpolation et extrapolation

L'observatoire de Mauna Loa (Hawaï) mesure depuis des décennies la concentration de dioxyde de carbone (CO_2) dans l'atmosphère, en parties par million (ppm). Le tableau donne quelques relevés ; on note x le nombre d'années écoulées depuis 2000 et y la concentration en ppm.

Année	2000	2005	2010	2015	2020	2024
Années écoulées x	0	5	10	15	20	24
Concentration y (ppm)	369,7	379,9	390,1	401,0	414,2	424,6

On saisit ces données dans un tableur et on demande une **courbe de tendance** linéaire. Le tableur affiche la droite d'équation $y = 2,28x + 368,5$, appelée **droite des moindres carrés**.



1. Que représente cette droite pour le nuage de points ? Faut-il connaître une formule pour l'obtenir ?
2. Interpréter le coefficient directeur de cette droite dans le contexte de l'exercice.
3. Aucun relevé n'a été fait en 2012. Estimer la concentration de CO_2 cette année-là à l'aide du modèle.
4. Estimer de la même façon la concentration de CO_2 en 2035.
5. Parmi les deux estimations précédentes, dire laquelle est une interpolation et laquelle est une extrapolation. Laquelle est la plus fiable ? Justifier.
6. **Question d'ouverture.** Que suppose-t-on sur les émissions mondiales de CO_2 lorsqu'on utilise cette extrapolation jusqu'en 2035 ?

Corrigé

1. La droite des moindres carrés est la droite qui passe « **au plus près** » de l'ensemble des points du nuage : parmi toutes les droites possibles, c'est celle qui rend globalement les plus petits possibles les écarts entre les points du nuage et la droite. Elle résume donc la tendance du nuage par une seule droite d'ajustement.

Aucune formule n'est à connaître pour l'obtenir : c'est le tableur (ou la calculatrice) qui la

calcule automatiquement et qui affiche son équation, ici $y = 2,28x + 368,5$.

2. Le coefficient directeur vaut 2,28 : lorsque x augmente de 1, y augmente de 2,28. Dans le contexte de l'exercice, cela signifie que, selon ce modèle, **la concentration de CO₂ dans l'atmosphère augmente d'environ 2,3 ppm par an** depuis 2000.

3. L'année 2012 correspond à $x = 2012 - 2000 = 12$ années écoulées. On remplace x par 12 dans l'équation de la droite :

$$y = 2,28 \times 12 + 368,5 = 27,36 + 368,5 = 395,86$$

On estime donc la concentration de CO₂ en 2012 à **environ 395,9 ppm**.

4. L'année 2035 correspond à $x = 2035 - 2000 = 35$. On calcule de la même façon :

$$y = 2,28 \times 35 + 368,5 = 79,8 + 368,5 = 448,3$$

Selon ce modèle, la concentration de CO₂ atteindrait **environ 448,3 ppm en 2035**.

5. Les relevés couvrent la plage allant de $x = 0$ (année 2000) à $x = 24$ (année 2024).

Pour 2012, on a $x = 12$, qui est **à l'intérieur** de cette plage : l'estimation de la question 3 est une **interpolation**.

Pour 2035, on a $x = 35$, qui est **en dehors** de cette plage : l'estimation de la question 4 est une **extrapolation**.

L'interpolation (celle de 2012) est la plus fiable : on estime une valeur encadrée par des relevés réels et proches (2010 et 2015), et le phénomène n'a aucune raison de s'être écarté brutalement de la tendance observée juste avant et juste après. Au contraire, l'estimation pour 2035 se situe 11 ans après le dernier relevé : rien ne garantit que la tendance observée se poursuive à l'identique aussi longtemps.

6. En prolongeant la droite jusqu'en 2035, on suppose que la concentration de CO₂ continuera d'augmenter **exactement au même rythme qu'entre 2000 et 2024**. Cela revient à supposer que les émissions mondiales de CO₂ resteront sur la même trajectoire : ni réduction importante liée à des politiques climatiques ambitieuses, ni au contraire accélération des émissions. C'est une hypothèse forte : l'évolution réelle dépendra des décisions collectives prises d'ici là. Autrement dit, l'avenir n'est pas déjà écrit dans la droite du tableur — c'est précisément l'enjeu des choix de société sur le climat.

Exercice 12 ★★ Interpoler ou extrapoler ?

Notions : Interpolation et extrapolation

La population d'une commune est relevée à chaque recensement. On note x le nombre d'années écoulées depuis 2006 et y le nombre d'habitants.

Année	2006	2011	2016	2021
Années écoulées x	0	5	10	15
Population y (habitants)	4 200	4 450	4 720	4 980

Le nuage de points étant presque rectiligne, on l'ajuste par la droite $\mathcal{D}$ d'équation :

$$y = 52x + 4\,190$$

1. Vérifier que la droite $\mathcal{D}$ est un ajustement raisonnable : calculer la valeur donnée par le modèle pour $x = 5$ et la comparer au relevé de 2011.
2. À l'aide du modèle, estimer la population de la commune pour chacune des années suivantes :
a. 2009 ; b. 2019 ; c. 2023 ; d. 2040.
3. Pour chacune de ces quatre estimations, dire s'il s'agit d'une **interpolation** ou d'une **extrapolation**. Justifier à l'aide de la plage des données.
4. Classer ces quatre estimations de la **plus fiable** à la **moins fiable**, en justifiant le classement.

Corrigé

1. Pour $x = 5$ (année 2011), le modèle donne :

$$y = 52 \times 5 + 4\,190 = 260 + 4\,190 = 4\,450$$

C'est exactement la population relevée en 2011 (4 450 habitants).

On peut compléter la vérification avec les autres relevés : pour $x = 0$, le modèle donne 4 190 habitants (relevé : 4 200) ; pour $x = 10$, il donne $52 \times 10 + 4\,190 = 4\,710$ (relevé : 4 720) ; pour $x = 15$, il donne $52 \times 15 + 4\,190 = 4\,970$ (relevé : 4 980). L'écart entre le modèle et les relevés ne dépasse jamais 10 habitants : la droite $\mathcal{D}$ est donc un ajustement tout à fait raisonnable du nuage.

2. Pour chaque année, on calcule d'abord le nombre x d'années écoulées depuis 2006, puis on remplace dans l'équation du modèle.

a. L'année 2009 correspond à $x = 2009 - 2006 = 3$, d'où $y = 52 \times 3 + 4\,190 = 156 + 4\,190 = 4\,346$. On estime la population de la commune en 2009 à environ 4 350 habitants.

b. L'année 2019 correspond à $x = 2019 - 2006 = 13$, d'où $y = 52 \times 13 + 4\,190 = 676 + 4\,190 = 4\,866$. On estime la population en 2019 à environ 4 870 habitants.

c. L'année 2023 correspond à $x = 2023 - 2006 = 17$, d'où $y = 52 \times 17 + 4\,190 = 884 + 4\,190 = 5\,074$. On estime la population en 2023 à environ 5 070 habitants.

d. L'année 2040 correspond à $x = 2040 - 2006 = 34$, d'où $y = 52 \times 34 + 4\,190 = 1\,768 + 4\,190 = 5\,958$. On estime la population en 2040 à environ 5 960 habitants.

3. Les données observées vont de $x = 0$ (recensement de 2006) à $x = 15$ (recensement de 2021). Pour 2009 ($x = 3$) et pour 2019 ($x = 13$) : la valeur de x est **à l'intérieur** de la plage des données. Ces deux estimations sont donc des **interpolations**.

Pour 2023 ($x = 17$) et pour 2040 ($x = 34$) : la valeur de x est **en dehors** de la plage des données. Ces deux estimations sont donc des **extrapolations**.

4. Classement de la plus fiable à la moins fiable :

- **Les deux interpolations, 2009 et 2019** (quasiment à égalité) : dans les deux cas, on estime une valeur encadrée par des recensements réels, et le nuage est très régulier sur toute la plage. Ce sont les estimations les plus sûres.
- **L'extrapolation proche, 2023** ($x = 17$, soit 2 ans seulement après le dernier recensement) : la tendance démographique n'a guère eu le temps de changer en si peu de temps, l'estimation reste donc plausible.
- **L'extrapolation lointaine, 2040** ($x = 34$, soit 19 ans après le dernier recensement) : en presque deux décennies, beaucoup d'événements peuvent modifier l'évolution de la population — construction d'un lotissement, fermeture d'une entreprise, évolution démographique générale. C'est de loin l'estimation la moins fiable.

On retrouve la règle générale : plus on s'éloigne de la plage des données observées, moins la prévision est fiable.

Exercice 13 ★★★ Vrai ou faux : ajustements et prévisions

Notions : Tableau croisé d'effectifs et fréquences, Ajustement affine (au jugé, Mayer, moindres carrés), Interpolation et extrapolation

Pour chacune des affirmations suivantes, dire si elle est **vraie** ou **fausse**. Chaque réponse doit être **justifiée** : par un argument court si l'affirmation est vraie, par un contre-exemple chiffré si elle est fausse.

- a. « Le point moyen d'un nuage de points est toujours l'un des points du nuage. »
- b. Dans un lycée de 100 élèves, on croise le caractère « fait du théâtre » avec le caractère « fille / garçon » :

	Théâtre	Pas de théâtre	Total
Filles	15	45	60
Garçons	5	35	40
Total	20	80	100

- « 75 % des élèves qui font du théâtre sont des filles, donc 75 % des filles font du théâtre. »
- c. « Si deux grandeurs évoluent ensemble (quand l'une augmente, l'autre augmente aussi), alors l'une est forcément la cause de l'autre. »
- d. « Une estimation faite à l'intérieur de la plage des données observées (interpolation) est en général plus fiable qu'une estimation faite au-delà de cette plage (extrapolation). »
- e. « La droite de Mayer d'un nuage de points passe par les deux points moyens partiels G_1 et G_2 . »

Corrigé

a. Faux. Le point moyen d'un nuage a pour coordonnées la moyenne $\bar{x}$ des abscisses et la moyenne $\bar{y}$ des ordonnées : c'est un point « de synthèse » du nuage, et rien n'oblige un point du nuage à se trouver exactement à cet endroit.

Contre-exemple chiffré : considérons le nuage formé des deux points (0 ; 0) et (4 ; 10). Son point moyen a pour coordonnées $\bar{x} = \frac{0+4}{2} = 2$ et $\bar{y} = \frac{0+10}{2} = 5$, c'est donc le point $G(2 ; 5)$. Or $G(2 ; 5)$ n'est aucun des deux points du nuage : l'affirmation est fausse.

b. Faux. Les deux pourcentages de l'affirmation ne sont pas calculés sur la même population de référence, et on ne peut pas les échanger.

La première fréquence est calculée **parmi les 20 élèves qui font du théâtre** : $\frac{15}{20} = 0,75$, donc 75 % des élèves de théâtre sont des filles — cette partie de la phrase est vraie. Mais la seconde fréquence doit être calculée **parmi les 60 filles** : $\frac{15}{60} = 0,25$, donc 25 % des filles font du théâtre, et non 75 %. Changer de population de référence change complètement le résultat : l'affirmation est fausse.

c. Faux. Deux grandeurs peuvent très bien évoluer ensemble sans que l'une soit la cause de l'autre : on parle alors seulement de **lien apparent**. Très souvent, un troisième facteur agit en même temps sur les deux grandeurs.

Exemple classique : chaque été, les ventes de glaces et le nombre de noyades augmentent ensemble. Les glaces ne provoquent évidemment pas les noyades : c'est la chaleur qui pousse à la fois à manger des glaces et à se baigner davantage, donc qui fait augmenter les deux grandeurs simultanément. Un lien apparent entre deux caractères ne prouve donc jamais, à lui seul, une causalité.

d. Vrai. Une interpolation estime une valeur située **entre** des valeurs réellement observées : sur cette plage, on sait comment le phénomène se comporte, puisque les données l'encadrent de

part et d'autre. Une extrapolation, au contraire, suppose que la tendance observée se poursuit **au-delà** de la plage des données, ce que rien ne garantit : le phénomène peut ralentir, saturer ou changer brutalement de comportement. C'est pourquoi, en général, une interpolation est plus fiable qu'une extrapolation — et une extrapolation est d'autant moins fiable qu'elle s'éloigne de la plage des données.

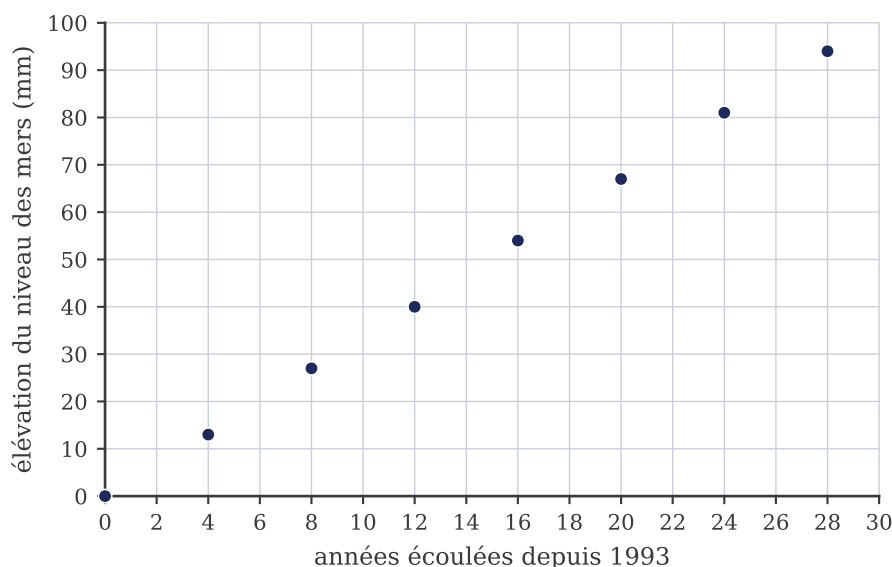
e. Vrai. C'est la définition même de la droite de Mayer : on partage le nuage, rangé dans l'ordre des abscisses croissantes, en deux moitiés ; on calcule le point moyen G_1 de la première moitié et le point moyen G_2 de la seconde ; la droite de Mayer est alors **la droite** (G_1G_2) . Elle passe donc par ces deux points **par construction**.

Exercice 14 ★★★ Synthèse : la montée du niveau des mers

Notions : Nuage de points, Point moyen, Ajustement affine (au jugé, Mayer, moindres carrés), Interpolation et extrapolation

Depuis 1993, des satellites mesurent l'élévation du niveau moyen des mers. Le tableau donne l'élévation y (en mm, par rapport au niveau de 1993) en fonction du nombre x d'années écoulées depuis 1993.

Années écoulées x	0	4	8	12	16	20	24	28
Élévation y (mm)	0	13	27	40	54	67	81	94



1. Décrire la forme du nuage de points. Un ajustement affine paraît-il adapté ?
2. On utilise la méthode de Mayer. Calculer les coordonnées du point moyen G_1 des 4 premiers points, puis celles du point moyen G_2 des 4 derniers points.
3. Déterminer l'équation réduite de la droite (G_1G_2) .
4. Interpréter le coefficient directeur de cette droite dans le contexte de l'exercice.
5. À l'aide de ce modèle, estimer l'élévation du niveau des mers en 2050 par rapport à 1993. Exprimer le résultat en mm puis en cm. S'agit-il d'une interpolation ou d'une extrapolation ?
6. **Question critique.** Les mesures les plus récentes montrent que la montée du niveau des mers **s'accélère** (elle est de plus en plus rapide). L'estimation de la question 5, obtenue avec un modèle affine, est-elle alors plutôt optimiste ou plutôt pessimiste ? Justifier.

Corrigé

1. Les huit points du nuage sont **presque alignés** : ils semblent se répartir le long d'une même droite. Le nuage est de plus **croissant** : quand x augmente, y augmente, et cette augmentation est très régulière (environ 13 à 14 mm tous les 4 ans). Un ajustement affine paraît donc tout à fait adapté pour modéliser cette évolution.
2. Point moyen G_1 des 4 premiers points (abscisses 0, 4, 8 et 12) : son abscisse est la moyenne des abscisses et son ordonnée est la moyenne des ordonnées correspondantes.

$$x_{G_1} = \frac{0 + 4 + 8 + 12}{4} = \frac{24}{4} = 6 \quad \text{et} \quad y_{G_1} = \frac{0 + 13 + 27 + 40}{4} = \frac{80}{4} = 20$$

Donc $G_1(6; 20)$.

Point moyen G_2 des 4 derniers points (abscisses 16, 20, 24 et 28) :

$$x_{G_2} = \frac{16 + 20 + 24 + 28}{4} = \frac{88}{4} = 22 \quad \text{et} \quad y_{G_2} = \frac{54 + 67 + 81 + 94}{4} = \frac{296}{4} = 74$$

Donc $G_2(22; 74)$.

3. L'équation réduite de la droite (G_1G_2) est de la forme $y = mx + p$.

Calcul du coefficient directeur :

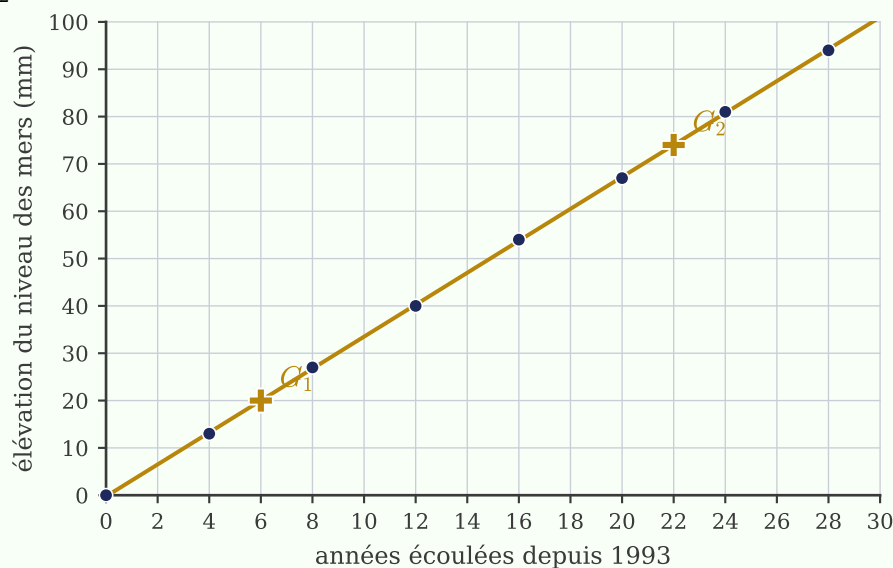
$$m = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{74 - 20}{22 - 6} = \frac{54}{16} = 3,375$$

Calcul de l'ordonnée à l'origine : la droite passe par $G_1(6; 20)$, donc les coordonnées de G_1 vérifient l'équation $y = 3,375x + p$:

$$20 = 3,375 \times 6 + p \Leftrightarrow 20 = 20,25 + p \Leftrightarrow p = 20 - 20,25 = -0,25$$

La droite de Mayer a donc pour équation réduite $y = 3,375x - 0,25$.

Vérification avec G_2 : $3,375 \times 22 - 0,25 = 74,25 - 0,25 = 74 = y_{G_2}$, donc la droite passe bien aussi par G_2 .



4. Le coefficient directeur vaut 3,375 : lorsque x augmente de 1, y augmente de 3,375. Dans le contexte de l'exercice, cela signifie que, selon ce modèle, **le niveau moyen des mers monte d'environ 3,4 mm par an** depuis 1993.

5. L'année 2050 correspond à $x = 2050 - 1993 = 57$ années écoulées. On remplace x par 57 dans l'équation de la droite de Mayer :

$$y = 3,375 \times 57 - 0,25 = 192,375 - 0,25 = 192,125$$

Le modèle prévoit donc une élévation d'environ 192 mm, c'est-à-dire **environ 19,2 cm**, du niveau moyen des mers en 2050 par rapport à 1993.

Les mesures s'arrêtent à $x = 28$ (année 2021) ; or $x = 57$ est très au-delà de cette plage : il s'agit d'une **extrapolation**, qui prolonge le modèle de presque 30 ans après les dernières données.

6. Le modèle affine utilisé suppose que le niveau des mers monte à un rythme **constant**, celui observé entre 1993 et 2021, soit environ 3,4 mm par an. Or les mesures les plus récentes montrent que la montée **s'accélère** : le rythme réel des prochaines décennies sera donc vraisemblablement

supérieur à 3,4 mm par an. L'élévation réelle en 2050 dépasserait alors les 19 cm environ calculés à la question 5.

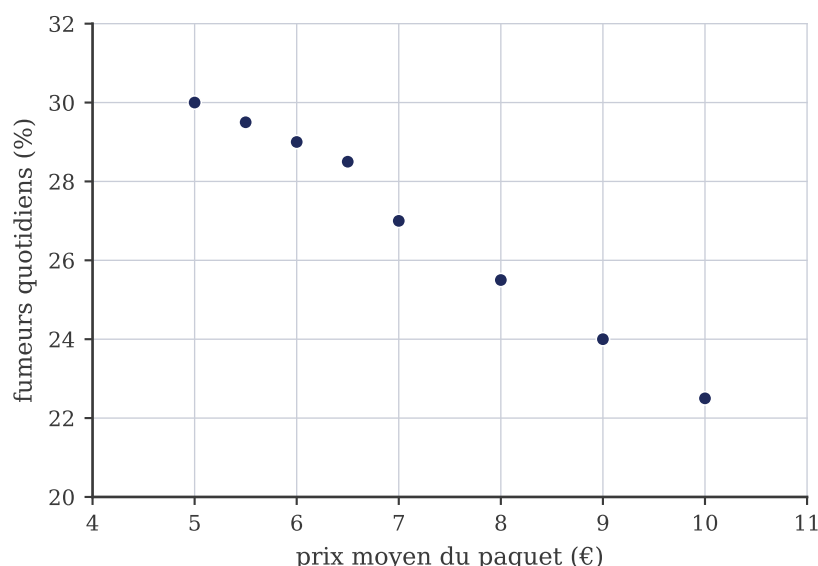
L'estimation du modèle affine est donc plutôt **optimiste** : c'est très probablement une **sous-estimation** de l'élévation réelle. Cet exemple montre qu'une extrapolation linéaire peut non seulement être imprécise, mais aussi se tromper dans un sens que l'on peut anticiper — et cette limite mathématique a ici des conséquences très concrètes pour les populations côtières.

Exercice 15 ★★★★★ Synthèse : prix du tabac et tabagisme

Notions : Nuage de points, Point moyen, Ajustement affine (au jugé, Mayer, moindres carrés), Interpolation et extrapolation

Dans un pays, on a relevé sur 8 années le prix moyen x du paquet de cigarettes (en euros) et la proportion y de fumeurs quotidiens dans la population adulte (en %).

Prix du paquet x (€)	5	5,5	6	6,5	7	8	9	10
Fumeurs quotidiens y (%)	30	29,5	29	28,5	27	25,5	24	22,5



Un tableur fournit la droite des moindres carrés associée à ce nuage : $y = -1,57x + 38,2$.

1. Décrire le lien apparent entre les deux caractères étudiés sur cette population.
2. Calculer les coordonnées du point moyen G du nuage.
3. Vérifier par le calcul que la droite des moindres carrés passe très près du point G .
4. Interpréter le coefficient directeur de la droite dans le contexte de l'exercice.
5. Estimer la proportion de fumeurs quotidiens pour un prix du paquet de 7,50 €. Interpolation ou extrapolation ?
6. Estimer de même la proportion de fumeurs pour un prix de 15 €. Pourquoi cette estimation doit-elle être considérée avec beaucoup de prudence ? Donner au moins deux raisons concrètes.
7. Résoudre l'inéquation $-1,57x + 38,2 < 20$. Interpréter le résultat, puis commenter la portée de cette prédiction.
8. **Esprit critique.** Ce lien apparent entre prix et tabagisme prouve-t-il que la hausse du prix **cause** la baisse du tabagisme ? Citer d'autres facteurs qui ont pu jouer sur la même période.

Corrigé

1. Le nuage de points est **décroissant** : quand le prix du paquet augmente, la proportion de fumeurs quotidiens diminue. Les points sont de plus en plus alignés, le long d'une droite qui descend. Il existe donc un **lien apparent** entre les deux caractères étudiés sur cette population :

plus le paquet est cher, plus la proportion de fumeurs quotidiens est faible.

2. L'abscisse du point moyen G est la moyenne des prix relevés :

$$\bar{x} = \frac{5 + 5,5 + 6 + 6,5 + 7 + 8 + 9 + 10}{8} = \frac{57}{8} = 7,125$$

Son ordonnée est la moyenne des proportions de fumeurs quotidiens :

$$\bar{y} = \frac{30 + 29,5 + 29 + 28,5 + 27 + 25,5 + 24 + 22,5}{8} = \frac{216}{8} = 27$$

Le point moyen du nuage est donc $G(7,125; 27)$.

3. On remplace x par l'abscisse de G , c'est-à-dire 7,125, dans l'équation de la droite des moindres carrés :

$$y = -1,57 \times 7,125 + 38,2 = -11,18625 + 38,2 = 27,01375$$

La droite donne environ 27,01 pour $x = 7,125$, alors que l'ordonnée de G vaut exactement 27 : l'écart n'est que d'environ 0,01. La droite des moindres carrés passe donc **très près** du point moyen G . (Elle passerait exactement par G si les coefficients affichés par le tableur n'étaient pas arrondis : c'est une propriété générale de cette droite.)

4. Le coefficient directeur vaut $-1,57$: il est négatif, ce qui traduit bien la décroissance du nuage. Concrètement, lorsque x augmente de 1, y diminue de 1,57 : selon ce modèle, **chaque euro d'augmentation du prix du paquet s'accompagne d'une baisse d'environ 1,6 point de pourcentage** de la proportion de fumeurs quotidiens.

5. Pour un paquet à 7,50 €, on remplace x par 7,5 :

$$y = -1,57 \times 7,5 + 38,2 = -11,775 + 38,2 = 26,425$$

On estime la proportion de fumeurs quotidiens à **environ** 26,4 % pour un prix du paquet de 7,50 €. Les prix observés vont de 5 € à 10 € ; comme 7,5 appartient à la plage $[5; 10]$, cette estimation est une **interpolation** : elle est fiable.

6. Pour un paquet à 15 €, on remplace x par 15 :

$$y = -1,57 \times 15 + 38,2 = -23,55 + 38,2 = 14,65$$

Le modèle prévoit **environ** 14,7 % de fumeurs quotidiens pour un paquet à 15 €. Mais 15 est très éloigné de la plage des données $[5; 10]$: c'est une **extrapolation**, à considérer avec beaucoup de prudence, pour au moins deux raisons concrètes.

Première raison : à un prix aussi élevé, une partie des fumeurs risque de se tourner vers le **marché parallèle** (contrebande, achats transfrontaliers) ; la proportion réelle de fumeurs baisserait alors beaucoup moins que prévu, même si les ventes officielles s'effondrent.

Deuxième raison : les **comportements ne sont pas linéaires**. Les fumeurs qui continuent malgré un prix élevé sont souvent les plus dépendants, donc les moins sensibles aux hausses de prix : la baisse pourrait ralentir bien avant d'atteindre le rythme prévu par la droite. Rien ne garantit que la tendance observée entre 5 € et 10 € se prolonge à l'identique jusqu'à 15 €.

7. On résout l'inéquation :

$$-1,57x + 38,2 < 20 \iff -1,57x < 20 - 38,2 \iff -1,57x < -18,2$$

On divise alors les deux membres par $-1,57$, qui est **négatif** : le sens de l'inégalité change.

$$x > \frac{-18,2}{-1,57} = \frac{18,2}{1,57} \approx 11,59$$

Interprétation : selon ce modèle, la proportion de fumeurs quotidiens passerait sous la barre des 20 % pour un prix du paquet supérieur à **environ** 11,60 €.

Portée de cette prédiction : un tel prix est **hors de la plage des données** (5 € à 10 €). La prédiction repose donc sur une extrapolation et hérite de toutes les réserves formulées à la question 6 : elle fournit un ordre de grandeur utile à la réflexion, mais ne constitue en aucun cas une certitude.

8. Non. Un lien apparent entre deux caractères ne prouve pas, à lui seul, que l'un **cause** l'autre. Sur la même période, de nombreux autres facteurs ont pu contribuer à la baisse du tabagisme : les **campagnes de prévention**, l'instauration du **paquet neutre**, l'**interdiction de fumer** dans les lieux publics, le remboursement des substituts nicotiniques, ou encore l'évolution générale des mentalités vis-à-vis du tabac. La hausse du prix est probablement **l'un des facteurs** de la baisse — les études des autorités de santé publique vont dans ce sens — mais le nuage de points, à lui seul, ne permet pas de conclure à une causalité : il montre seulement que les deux évolutions se sont produites en même temps.