Exercice 1 — Le sens des moindres carrés

4 points

1) $\bar{x} = \dfrac{1+2+3}{3} = 2$ et $\bar{y} = \dfrac{2+3+5}{3} = \dfrac{10}{3}$. Le point moyen est $\boxed{G\left(2\,;\dfrac{10}{3}\right)}$.

2) On calcule chaque écart vertical :

$$2-(1{,}5+b) = 0{,}5-b, \qquad 3-(3+b) = -b, \qquad 5-(4{,}5+b) = 0{,}5-b.$$

Donc, pour tout réel b :

$$S(b) = (0{,}5-b)^2 + (-b)^2 + (0{,}5-b)^2 = 2\left(0{,}25 - b + b^2\right) + b^2, \quad \text{soit} \quad \boxed{S(b) = 3b^2 - 2b + 0{,}5}.$$

3) S est une fonction trinôme de coefficient dominant $3 > 0$: elle admet un minimum, atteint en $b = -\dfrac{-2}{2\times 3} = \dfrac{1}{3}$.

Ce minimum vaut $S\left(\dfrac{1}{3}\right) = 3\times\dfrac{1}{9} - \dfrac{2}{3} + \dfrac{1}{2} = \dfrac{1}{3} - \dfrac{2}{3} + \dfrac{1}{2} = \dfrac{2-4+3}{6} = \dfrac{1}{6}$.

Ainsi $\boxed{b = \dfrac{1}{3}}$ et $\boxed{S\left(\dfrac{1}{3}\right) = \dfrac{1}{6}}$.

La droite « des moindres carrés » parmi celles de pente $1{,}5$ a donc pour équation $y = 1{,}5x + \dfrac{1}{3}$.

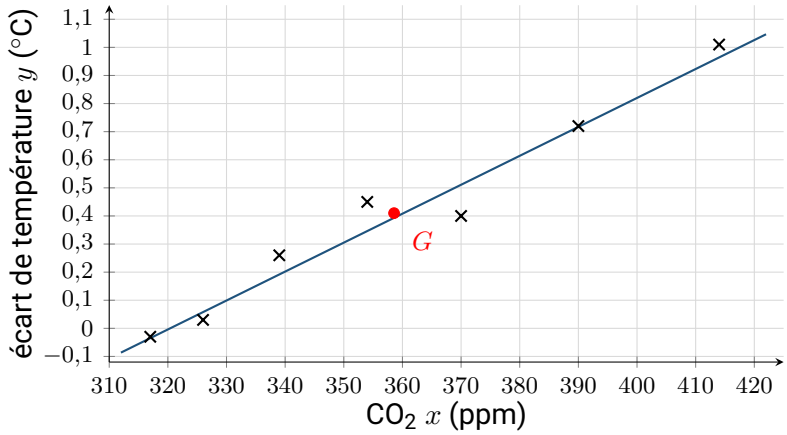
4) Pour $x = 2$: $1{,}5\times 2 + \dfrac{1}{3} = 3 + \dfrac{1}{3} = \dfrac{10}{3} = \bar{y}$. La droite passe bien par G, comme le prévoit le cours pour la droite de régression.

Exercice 2 — Le climat se réchauffe

9 points

Partie A — Un ajustement affine

1) Nuage de points (la droite de régression de la question 3 est tracée) :



2) $\bar{x} = \dfrac{317+326+339+354+370+390+414}{7} = \dfrac{2\,510}{7} \approx 358{,}57$

et $\bar{y} = \dfrac{-0{,}03+0{,}03+0{,}26+0{,}45+0{,}40+0{,}72+1{,}01}{7} = \dfrac{2{,}84}{7} \approx 0{,}41.$

Le point moyen est $\boxed{G(358{,}57\,;0{,}41)}$.

3) **a.** La calculatrice (mode statistiques à deux variables, régression linéaire) donne $a \approx 0{,}0103$, $b \approx -3{,}30$ et $r \approx 0{,}982$:

$$\boxed{y = 0{,}0103\,x - 3{,}30} \qquad \text{et} \qquad \boxed{r \approx 0{,}982}$$

b. r est proche de 1 : la corrélation linéaire entre x et y est forte et positive ; un ajustement affine est pertinent, et l'écart de température augmente avec la concentration de CO_2 (d'environ $0{,}01\,°C$ par ppm supplémentaire).

4) **a.** Pour $x = 400$: $y = 0{,}0103 \times 400 - 3{,}30 = 4{,}12 - 3{,}30 = 0{,}82$. Le modèle donne environ $0{,}8\,°C$. L'écart observé en 2015, $0{,}90\,°C$, est un peu supérieur mais du même ordre : le modèle rend bien compte des données dans la plage étudiée (l'écart de $0{,}08\,°C$ est comparable à la dispersion des points autour de la droite, par exemple $0{,}40$ en 2000 contre $0{,}45$ en 1990).

b. Pour $x = 500$: $y = 0{,}0103 \times 500 - 3{,}30 = 5{,}15 - 3{,}30 = 1{,}85$. Le modèle donne environ $1{,}9\,°C$ d'écart vers 2050.

Partie B — Regard critique

5) 400 ppm est compris entre 317 et 414, les valeurs extrêmes observées : la question 4.a est une **interpolation**. 500 ppm est en dehors de cette plage : la question 4.b est une **extrapolation**. L'extrapolation est la plus fragile : rien ne garantit que la relation reste affine au-delà des valeurs observées (le système climatique peut réagir plus fortement, ou moins, avec des effets d'inertie et de seuils), et les autres facteurs peuvent changer d'ici 2050.

6) L'élève a raison sur le principe : une corrélation, même forte, ne démontre pas une causalité. Deux grandeurs qui augmentent régulièrement avec le temps (population mondiale, CO_2, température…) sont automatiquement corrélées entre elles, le temps jouant le rôle de variable cachée ; le coefficient r ne dit pas laquelle agit sur l'autre.
Mais la causalité du CO_2 n'est pas établie par la statistique seule : elle repose sur la **physique** de l'effet de serre (le CO_2 absorbe le rayonnement infrarouge émis par la Terre, mécanisme connu et mesuré en laboratoire depuis le XIX^{e} siècle), sur des modèles qui reproduisent les températures observées seulement en tenant compte des gaz à effet de serre, et sur des observations indépendantes (refroidissement de la haute atmosphère, par exemple). La corrélation est *cohérente* avec ce mécanisme ; c'est le mécanisme qui fonde la causalité, pas r.

Exercice 3 — Du chocolat aux prix Nobel

7 points

1) $\bar{x} = \dfrac{11{,}5 + 11 + 6{,}5 + 6 + 4{,}5 + 3{,}5 + 3 + 2}{8} = \dfrac{48}{8} = 6$

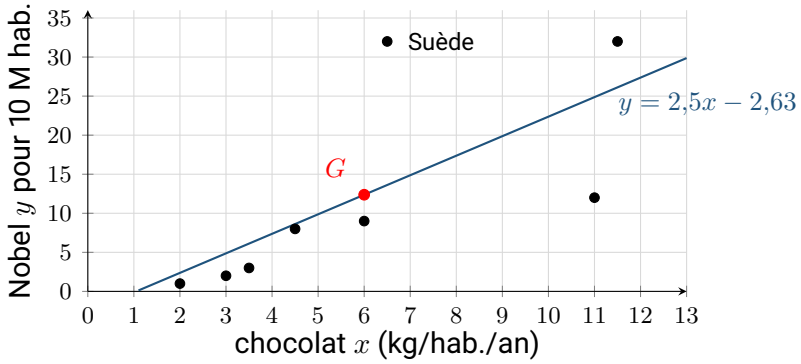
et $\bar{y} = \dfrac{32 + 12 + 32 + 9 + 8 + 3 + 2 + 1}{8} = \dfrac{99}{8} = 12{,}375.$

Le point moyen est $\boxed{G(6\,;12{,}4)}$.

2) La calculatrice donne $a = 2{,}5$, $b \approx -2{,}63$ et $r \approx 0{,}70$:

$$\boxed{y = 2{,}5\,x - 2{,}63} \qquad \text{et} \qquad \boxed{r \approx 0{,}70}$$

La corrélation est positive et assez nette, sans être très forte (les points sont dispersés autour de la droite).



3) **a.** Pour $x = 9$: $y = 2{,}5 \times 9 - 2{,}63 = 22{,}5 - 2{,}63 = 19{,}87$. Le modèle prévoit $\boxed{\text{environ } 20 \text{ prix Nobel pour } 10 \text{ millions d'habitants}}$.
9 est compris entre 2 et $11{,}5$, les consommations extrêmes observées : c'est une **interpolation**.

b. Pour $x = 20$: $y = 2{,}5 \times 20 - 2{,}63 = 47{,}37$,
soit $\boxed{\text{environ } 47 \text{ prix Nobel pour } 10 \text{ millions d'habitants}}$.
C'est une **extrapolation** ($20 > 11{,}5$) : aucun pays n'approche une telle consommation, et rien n'indique que le lien affine se prolonge ; cette valeur n'a guère de sens.

4) Sans la Suède (sept pays), la calculatrice donne $\boxed{r' \approx 0{,}86}$ (et $y \approx 2{,}38x - 4{,}55$).
$r' > r$: $\boxed{\text{le lien affine paraît plus fort}}$ une fois retiré ce point atypique (la Suède compte beaucoup de prix Nobel pour une consommation moyenne, notamment parce que les prix y sont décernés). Un seul point peut donc modifier sensiblement le coefficient de corrélation d'un petit nuage.

5) La corrélation positive signifie seulement que les pays où l'on mange plus de chocolat comptent, en moyenne, plus de lauréats : $\boxed{\text{elle ne prouve pas que le chocolat soit la cause}}$. Une **variable cachée** explique bien mieux le lien : la **richesse du pays** (PIB par habitant). Les pays riches consomment davantage de produits comme le chocolat *et* financent davantage la recherche et l'enseignement supérieur, d'où plus de prix Nobel. Les deux variables dépendent d'une même troisième, sans agir l'une sur l'autre.