Terminale maths complémentaires / Statistiques à deux variables — corrigé

benprof particulier.fr

1. *Rappels : moyenne, variance, écart type*

Ex. 1 ★☆☆ **Automatismes : moyenne, variance, écart type**

(a) $\overline{x} = \frac{12+15+9+14}{4} = \frac{50}{4} = 12{,}5$.

(b) $\overline{x} = \frac{24}{4} = 6$. Écarts : -3 ; -1 ; 1 ; 3 (somme nulle). Carrés : 9 ; 1 ; 1 ; 9. $V(x) = \frac{20}{4} = 5$.

(c) König-Huygens : $V(x) = \frac{520}{10} - 7^2 = 52 - 49 = 3$.

(d) $\sigma = \sqrt{2{,}25} = 1{,}5$ cm : l'écart type est dans l'unité des données, la variance en cm².

(e) Nouvelle moyenne : $20 + 5 = 25$. L'écart type ne change pas : il vaut toujours 3, car ajouter une constante décale les valeurs sans les disperser.

(f) $\overline{y} = -0{,}5 \times 4 + 10 = 8$ et $\sigma_y = |-0{,}5| \times 1{,}5 = 0{,}75$ (un écart type n'est jamais négatif).

Ex. 2 ★☆☆ **Automatismes : vrai ou faux ?**

(a) Faux. La moyenne de 4, 16 et 36 vaut $\frac{56}{3}$: c'est la moyenne des carrés. Avec $\overline{x} = 4$, les écarts sont -2 ; 0 ; 2, donc $V = \frac{4+0+4}{3} = \frac{8}{3}$. König-Huygens donne bien $\frac{56}{3} - 16 = \frac{8}{3}$.

(b) Faux. La variance est une moyenne de carrés d'écarts : elle s'exprime en kg². C'est l'écart type qui est en kg.

(c) Vrai. D'après König-Huygens, la moyenne des carrés moins le carré de la moyenne vaut $V(x)$. Or $V(x) \geqslant 0$ comme moyenne de carrés.

(d) Faux. Si $y_i = 3x_i$, alors $\sigma_y = 3\sigma_x$, donc $V(y) = \sigma_y^2 = 9\,\sigma_x^2 = 9\,V(x)$: la variance est multipliée par $3^2 = 9$. C'est l'écart type qui est multiplié par 3.

(e) Faux. Les séries 9 ; 11 et 0 ; 20 ont toutes deux pour moyenne 10. Leurs écarts types valent $\sqrt{\frac{1+1}{2}} = 1$ et $\sqrt{\frac{100+100}{2}} = 10$. La moyenne ne renseigne pas sur la dispersion.

(f) Vrai. Si $\sigma = 0$, alors $V = 0$: une somme de carrés nulle impose que chaque écart $x_i - \overline{x}$ soit nul. Toutes les valeurs sont donc égales à $\overline{x}$.

Ex. 3 ★★☆ **Une série avec effectifs**

(a) $\overline{x} = \frac{0\times1+1\times4+2\times6+3\times5+4\times4}{20} = \frac{47}{20} = 2{,}35$ livres.

(b) Amine a oublié les effectifs : il compte chaque valeur une seule fois. Or 2, 3 et 4 livres concernent beaucoup plus d'élèves que 0 livre. Il faut pondérer chaque valeur par son effectif et diviser par $N = 20$.

(c) $\sum n_i x_i^2 = 1 \times 0 + 4 \times 1 + 6 \times 4 + 5 \times 9 + 4 \times 16 = 137$.

(d) $V(x) = \frac{137}{20} - 2{,}35^2 = 6{,}85 - 5{,}5225 = 1{,}3275$.

(e) $\sigma_x = \sqrt{1{,}3275} \approx 1{,}15$ livre.

Ex. 4 ★★★ **Trois erreurs d'élèves**

(a) $\overline{x} = \frac{760}{5} = 152$ g. Écarts : 0 ; -4 ; 8 ; -7 ; 3 (somme nulle). Carrés : 0 ; 16 ; 64 ; 49 ; 9, de somme 138. Donc $V = \frac{138}{5} = 27{,}6$ g².

(b) Lina trouve $\frac{138}{4} = 34{,}5$. Elle divise par $n - 1$: c'est le s_x de la calculatrice. Le cours divise par $n = 5$, ce qui donne $V = 27{,}6$ g².

(c) Nolan obtient $\frac{115\,658 - 23\,104}{5} = 18\,510{,}8$: c'est absurde. Seule la somme des carrés se divise par 5 : $V = \frac{115\,658}{5} - 152^2 = 23\,131{,}6 - 23\,104 = 27{,}6$ g².

(d) Sami arrondit 27,6 à 28 et annonce $\sqrt{28} \approx 5{,}29$ g. Or $\sigma = \sqrt{27{,}6} \approx 5{,}25$ g : il a arrondi trop tôt. On garde la valeur exacte de V et on arrondit seulement à la fin. L'écart type des masses vaut environ 5,25 g.

Ex. 5 ★★★ **Températures : changer d'unité pour comparer**

(a) $\overline{C} = \frac{65}{5} = 13$ °C. Écarts : -1 ; 2 ; 1 ; -2 ; 0, carrés de somme 10.
Donc $V = \frac{10}{5} = 2$ et $\sigma = \sqrt{2} \approx 1{,}41$ °C.

(b) Changement affine $F = 1{,}8C + 32$: la moyenne vaut $1{,}8 \times 13 + 32 = 55{,}4$ °F. L'écart type vaut $1{,}8\sqrt{2} \approx 2{,}55$ °F, car 32 ne change pas la dispersion.

(c) Inès trouve environ $-15{,}28$: un écart type négatif est impossible. On a $C = \frac{F-32}{1{,}8}$, donc $\sigma_C = \frac{4{,}5}{1{,}8} = 2{,}5$ °C (le -32 ne joue pas). Moyenne : $\frac{57{,}2-32}{1{,}8} = 14$ °C.

(d) Clermont a été plus chaude en moyenne (14 °C contre 13 °C), et ses températures ont été plus dispersées (2,5 °C contre environ 1,41 °C). Annecy a donc eu un temps plus frais mais plus régulier.

Ex. 6 ★★★ **La moyenne, valeur la plus proche de toute la série**

(a) On a $\sum x_i = 4+7+5+9+10 = 35$ et $\sum x_i^2 = 16+49+25+81+100 = 271$. En développant chaque carré, $(x_i - t)^2 = x_i^2 - 2x_i t + t^2$, donc $\sum (x_i - t)^2 = 271 - 70t + 5t^2$. Ainsi $f(t) = t^2 - 14t + 54{,}2$.

(b) f est un trinôme avec $\alpha = 1 > 0$: son minimum est atteint en $t = -\frac{-14}{2\times 1} = 7$. Il vaut $f(7) = 49 - 98 + 54{,}2 = 5{,}2$. Or $\overline{x} = \frac{35}{5} = 7$ min, et avec les écarts -3 ; 0 ; -2 ; 2 ; 3, on obtient $V(x) = \frac{9+0+4+4+9}{5} = 5{,}2$. Le minimum est donc atteint pour la moyenne, et il est égal à la variance.

(c) $f(0) = \frac{271}{5} = 54{,}2$ est la moyenne des carrés des temps, et $f(0) - f(7) = 54{,}2 - 5{,}2 = 49 = 7^2 = \overline{x}^2$.
Donc $V(x) = \frac{1}{5}\sum x_i^2 - \overline{x}^2$: c'est la formule de König-Huygens. On voit aussi que la moyenne des carrés ($54{,}2$) n'est pas le carré de la moyenne (49).

(d) Non : $f(0)$ est la moyenne des carrés, pas la variance. Il faut prendre la moyenne des carrés des écarts, $V(x) = f(7) = 5{,}2$.
Donc $\sigma_x = \sqrt{5{,}2} \approx 2{,}28$ min, alors que $\sqrt{54{,}2} \approx 7{,}36$ ne mesure pas la dispersion. L'écart type des temps d'attente est d'environ 2,28 min.

(e) $\sum (x_i - t)^2 = \sum x_i^2 - 2t\sum x_i + nt^2$ et $\sum x_i = n\overline{x}$, donc $f(t) = t^2 - 2\overline{x}\,t + \frac{1}{n}\sum x_i^2$. C'est un trinôme avec $\alpha = 1 > 0$, minimal en $t = \frac{2\overline{x}}{2} = \overline{x}$. Par définition, $f(\overline{x}) = V(x)$, et $V(x) \geqslant 0$ car c'est une moyenne de carrés. Or $f(\overline{x}) = \overline{x}^2 - 2\overline{x}^2 + \frac{1}{n}\sum x_i^2 = \frac{1}{n}\sum x_i^2 - \overline{x}^2$. On en déduit $\frac{1}{n}\sum x_i^2 \geqslant \overline{x}^2$, avec égalité seulement si toutes les valeurs sont égales.

Ex. 7 ★★★ **Portions de fromage : deux lignes de conditionnement**

(a) Chaque valeur est pondérée par son effectif : $\overline{x} = \frac{2\times(-4)+5\times(-2)+8\times 0+4\times 2+1\times 4}{20} = \frac{-6}{20} = -0{,}3$. Comme $m = x + 250$, on a $\overline{m} = \overline{x} + 250 = 249{,}7$ g. La masse moyenne d'une portion de la ligne A est 249,7 g.

(b) $\frac{1}{20}\sum n_i x_i^2 = \frac{2\times 16+5\times 4+8\times 0+4\times 4+1\times 16}{20} = \frac{84}{20} = 4{,}2$, donc $V(x) = 4{,}2 - (-0{,}3)^2 = 4{,}2 - 0{,}09 = 4{,}11$. Sans parenthèses, la calculatrice lit $-0{,}3^2$ comme $-(0{,}3^2) = -0{,}09$. Elle donne alors $4{,}2 + 0{,}09 = 4{,}29$, ce qui est faux. Il faut taper $(-0{,}3)^2$.

Ex. 7 (suite)

(c) $m = x + 250$: ajouter une constante décale toutes les valeurs sans changer les écarts à la moyenne.
Donc $V(m) = V(x) = 4{,}11$ g^2 et $\sigma_m = \sqrt{4{,}11} \approx 2{,}03$ g. L'écart type des masses de la ligne A est d'environ 2,03 g. La variance, elle, est en g^2.

(d) $d = -m + 250$ est un changement affine avec $a = -1$ et $b = 250$.
Donc $\overline{d} = -\overline{m} + 250 = 250 - 249{,}7 = 0{,}3$ g et $\sigma_d = |-1| \times \sigma_m \approx 2{,}03$ g. Un écart type n'est jamais négatif. En moyenne, il manque 0,3 g par portion, et la dispersion ne change pas.

(e) Ligne A : $249{,}7 \in [249;251]$ et $2{,}03 \leqslant 2{,}5$, elle est conforme. Ligne B : $250{,}1 \in [249;251]$, mais $3{,}4 > 2{,}5$, elle n'est pas conforme. La ligne B a la moyenne la plus proche de 250 g, mais ses masses sont trop dispersées. Comparer les moyennes ne suffit pas : seule la ligne A respecte la norme.

2. *Nuage de points et point moyen*

Ex. 8 ★☆☆ **Automatismes : le point moyen d'une série**

(a) $\overline{x} = \frac{14+18+21+25+27}{5} = \frac{105}{5} = 21$: la température moyenne est de 21 °C.

(b) $\overline{y} = \frac{32+45+51+68+74}{5} = \frac{270}{5} = 54$: en moyenne, le glacier vend 54 cornets par jour.

(c) $G(21\,;54)$.

(d) Les valeurs sont déjà rangées : la médiane des x_i vaut 21 et celle des y_i vaut 51. Le point des médianes est $(21\,;51)$. Ce n'est pas G, car $\overline{y} = 54 \neq 51$: G se calcule avec les moyennes, pas avec les médianes.

(e) Non. Le seul point de la série d'abscisse 21 est $(21\,;51)$, et son ordonnée n'est pas 54. G est un point construit, qui ne fait pas partie de la série.

(f) Les x_i ne changent pas, donc $\overline{x} = 21$. Seule l'ordonnée change : $\overline{y} = \frac{32+45+51+68+84}{5} = \frac{280}{5} = 56$. Le nouveau point moyen est $G(21\,;56)$: il monte de $\frac{10}{5} = 2$ cornets.

Ex. 9 ★☆☆ **Automatismes : vrai ou faux sur le nuage**

(a) Faux. Un nuage allongé peut être courbé : les points $(1\,;1)$, $(2\,;4)$, $(3\,;9)$, $(4\,;16)$ montent de 3, puis 5, puis 7. Même presque aligné, un nuage suggère un ajustement affine, il ne le prouve pas.

Ex. 9 **(suite)**

(b) Faux. L'axe commence à 480 et non à 0 : à l'écran, les hauteurs valent $492-480=12$ et $510-480=30$, d'où l'impression d'un rapport de 2,5. En réalité, $\frac{510-492}{492}\approx 0{,}037$: y n'a augmenté que d'environ 3,7 %.

(c) Faux. $\overline{x}=\frac{2+4+6}{3}=4$, mais $\overline{y}=\frac{10+14+30}{3}=18$.
Donc $G(4\,;18)$, pas le point du milieu du tableau.

(d) Vrai. Les n points ont pour somme des abscisses $n\overline{x}$. On ajoute $\overline{x}$, ce qui donne $(n+1)\overline{x}$. La nouvelle moyenne vaut donc encore $\overline{x}$. Le raisonnement est le même pour $\overline{y}$.

(e) Vrai. La moyenne des y_i+3 vaut $\overline{y}+3$, et $\overline{x}$ ne change pas : G monte de 3.

(f) Faux. 2026 vient 6 ans après 2020, donc $x=1+6=7$.

Ex. 10 ★★☆ **Autopartage à Grenoble : lire un nuage**

(a) 2022 correspond à $x=3$, car $x=0$ pour 2019. On lit $y=4{,}8$: en 2022, le service comptait environ 4 800 abonnés.

(b) Les variations annuelles valent $-1{,}2$; $+2{,}5$; $+0{,}7$; $+0{,}6$; $+0{,}7$. Le nuage est croissant dans l'ensemble, avec un point isolé : celui de 2020, $(1\,;1{,}6)$, nettement sous la tendance. Hors de ce point, la hausse est régulière : $+1{,}3$ en deux ans de 2019 à 2021, puis environ $+0{,}7$ par an. Ces points sont presque alignés. Un ajustement affine paraît donc adapté, à condition de signaler le point de 2020, ou de l'écarter.

(c) $\overline{x}=\frac{0+1+2+3+4+5}{6}=2{,}5$ et $\overline{y}=\frac{2{,}8+1{,}6+4{,}1+4{,}8+5{,}4+6{,}1}{6}=\frac{24{,}8}{6}\approx 4{,}13$.
Donc $G(2{,}5\,;4{,}13)$.

(d) Non. L'axe des ordonnées commence à 1, pas à 0 : à l'écran, les hauteurs valent $1{,}6-1=0{,}6$ et $6{,}1-1=5{,}1$, d'où un rapport de 8,5. Le vrai rapport est $\frac{6{,}1}{1{,}6}\approx 3{,}81$: entre 2020 et 2024, le nombre d'abonnés a été multiplié par environ 3,8.

Ex. 11 ★★☆ **Course à pied à Annecy : corrige l'élève**

(a) On place en abscisse la variable explicative : l'entraînement x, que le coureur choisit. Le temps sur 10 km, la variable expliquée, va en ordonnée, quel que soit l'ordre des colonnes.

(b) Nolan a pris un point de la série, celui du milieu du tableau. Or G se calcule avec deux moyennes séparées : $\overline{x}=\frac{3+5+6+8+9}{5}=\frac{31}{5}=6{,}2$ et $\overline{y}=\frac{58+52+50+46+43}{5}=\frac{249}{5}=49{,}8$.
Donc $G(6{,}2\,;49{,}8)$.

Ex. 11 **(suite)**

(c) Quand le nombre d'heures d'entraînement augmente, le temps sur 10 km diminue : dans ce club, les coureurs qui s'entraînent le plus courent le plus vite.

(d) Non. B a pour coordonnées $(5\,;52)$, avec $5<6{,}2$ et $52>49{,}8$: son point est à gauche de G et au-dessus. Il s'entraîne donc moins que la moyenne et met plus de temps que la moyenne. Ici, être « au-dessus » signifie courir plus lentement : B court moins bien que la moyenne du club.

Ex. 12 ★★☆ **Engrais près de Chartres : sous-groupes et point ajouté**

(a) $\overline{x}=\frac{40+60+80+100+120+140}{6}=\frac{540}{6}=90$ et $\overline{y}=\frac{52+58+64+69+73+77}{6}=\frac{393}{6}=65{,}5$.
Donc $G(90\,;65{,}5)$.

(b) $G_1\left(\frac{40+60+80}{3}\,;\frac{52+58+64}{3}\right)=(60\,;58)$ et $G_2\left(\frac{100+120+140}{3}\,;\frac{69+73+77}{3}\right)=(120\,;73)$.

(c) Les deux coordonnées changent : $\overline{x}'=\frac{540+160}{7}=100$ et $\overline{y}'=\frac{393+74}{7}=\frac{467}{7}\approx 66{,}71$.
Donc $G'(100\,;66{,}71)$.

(d) Les variations sont $+6$, $+6$, $+5$, $+4$, $+4$, puis -3. La hausse ralentit, puis le rendement baisse : le nuage est courbé. Un ajustement affine convient mal, surtout pour les fortes doses d'engrais.

Ex. 13 ★★★ **Application nantaise : un nuage qui se courbe**

(a) Les écarts successifs valent 0,5 ; 0,7 ; 0,8 ; 1,1 ; 1,3 et 1,7. Le nuage est croissant et ne présente aucun point isolé. Mais les écarts augmentent : la croissance s'accélère et le nuage se courbe vers le haut. Un ajustement affine n'est donc pas adapté.

(b) Les quotients valent environ 1,24 ; 1,27 ; 1,24 ; 1,27 ; 1,25 ; 1,26. Ils sont presque constants, voisins de 1,25 : chaque trimestre, le nombre d'utilisateurs est multiplié par environ 1,25, soit une hausse d'environ 25 %. Cela suggère un modèle exponentiel $y=k\,e^{ax}$, qu'on ramène à un ajustement affine par le changement de variable $z=\ln y$.

(c) $\overline{x}=\frac{1+2+3+4+5+6+7}{7}=4$ et $\overline{y}=\frac{2{,}1+2{,}6+3{,}3+4{,}1+5{,}2+6{,}5+8{,}2}{7}=\frac{32}{7}\approx 4{,}57$, donc $G(4\,;4{,}57)$.
Le point du 4e trimestre est $(4\,;4{,}1)$: il a la bonne abscisse, mais pas la bonne ordonnée. Comme la croissance s'accélère, les dernières valeurs tirent la moyenne vers le haut. G n'est donc pas un point de la série, et Léa a tort.

Ex. 13 **(suite)**

(d) $x = 1$ correspond au 1er trimestre 2024, donc $x = 8$ correspond au 4e trimestre 2025. On cherche y_8 tel que $\frac{32+y_8}{8} = 5{,}2$, soit $y_8 = 41{,}6 - 32 = 9{,}6$. L'abscisse change elle aussi : $\overline{x} = \frac{36}{8} = 4{,}5$. Il faut donc atteindre 9 600 utilisateurs au 4e trimestre 2025, et le nouveau point moyen est $G'(4{,}5\,;5{,}2)$.

Ex. 14 ★★★ **Vélos électriques d'occasion à Rennes : sous-groupes et point ajouté**

(a) Plus le vélo est âgé, plus son prix baisse : le nuage est décroissant. Il est allongé et ne présente aucun point isolé, mais seul un calcul pourra confirmer qu'un ajustement affine convient. $\overline{x} = \frac{1+1+2+3+4+5+6+8}{8} = \frac{30}{8} = 3{,}75$ et $\overline{y} = \frac{1\,900+1\,750+1\,550+1\,400+1\,150+1\,000+850+600}{8} = \frac{10\,200}{8} = 1\,275$, donc $G(3{,}75\,;1\,275)$. Aucun vélo n'a ces caractéristiques : G n'est pas un point de la série.

(b) Le milieu vaut $\left(\frac{1+8}{2}\,;\frac{1\,900+600}{2}\right) = (4{,}5\,;1\,250)$, alors que $G(3{,}75\,;1\,275)$. Ce milieu ne tient compte que de deux vélos sur huit. Or les vélos récents sont nombreux, ce qui fait baisser l'âge moyen et monter le prix moyen. Le point moyen se calcule avec toutes les valeurs : Malo a tort.

(c) G_1 : $\frac{1+1+2+3}{4} = 1{,}75$ et $\frac{1\,900+1\,750+1\,550+1\,400}{4} = 1\,650$, donc $G_1(1{,}75\,;1\,650)$. G_2 : $\frac{4+5+6+8}{4} = 5{,}75$ et $\frac{1\,150+1\,000+850+600}{4} = 900$, donc $G_2(5{,}75\,;900)$. Le milieu de $[G_1G_2]$ est $\left(\frac{1{,}75+5{,}75}{2}\,;\frac{1\,650+900}{2}\right) = (3{,}75\,;1\,275)$, c'est-à-dire G. Cela vient de ce que les deux groupes ont le même effectif (4 vélos chacun) : la moyenne générale est alors la moyenne des deux moyennes, sur chaque coordonnée.

(d) L'ordonnée reste égale à 1 275 si $\frac{10\,200+p}{9} = 1\,275$, soit $p = 11\,475 - 10\,200 = 1\,275$ €. L'abscisse devient $\frac{40}{9} \approx 4{,}44$. Ce prix n'est pas vraisemblable : un vélo de 10 ans vendu 1 275 € coûterait plus cher qu'un vélo de 4 ans (1 150 €). Ce serait un point isolé, contraire à la tendance décroissante.

(e) $\overline{x} = \frac{30+10}{9} = \frac{40}{9} \approx 4{,}44$ et $\overline{y} = \frac{10\,200+450}{9} = \frac{10\,650}{9} \approx 1\,183{,}33$.
Donc $G'(4{,}44\,;1\,183{,}33)$: l'ajout de ce vélo modifie les deux coordonnées du point moyen.

3. *Droite des moindres carrés*

Ex. 15 ★☆☆ **Automatismes : a et b à partir des indicateurs**

(a) $a = \frac{6}{2{,}5} = 2{,}4$ et $b = 20 - 2{,}4 \times 4 = 10{,}4$: $y = 2{,}4x + 10{,}4$.

(b) $b = \overline{y} - a\overline{x} = 14 - (-3) \times 5 = 14 + 15 = 29$: $y = -3x + 29$.

(c) $V(x) = \sigma_x^2 = 4$, donc $a = \frac{10}{4} = 2{,}5$ (et non $\frac{10}{2}$). Puis $b = 40 - 2{,}5 \times 3 = 32{,}5$.

(d) Non : $a = \frac{\text{cov}(x,y)}{V(x)} = \frac{5}{4} = 1{,}25$. Le quotient $\frac{V(y)}{V(x)} = 2{,}25$ ne donne pas le coefficient directeur.

(e) $b = 10 - \frac{7}{3} \times 2 = \frac{16}{3} \approx 5{,}33$. Avec a arrondi : $10 - 2{,}33 \times 2 = 5{,}34$. On garde donc a non arrondi pour calculer b.

(f) Faux : $1{,}5 \times 6 + 2 = 11 \neq 10$. La droite ne passe pas par G, donc ce n'est pas la droite des moindres carrés.

Ex. 16 ★☆☆ **Automatismes : résidus et estimations**

(a) $e = 14 - (2 \times 5 + 3) = 14 - 13 = 1$.

(b) $e = 9 - (2 \times 4 + 3) = 9 - 11 = -2$: le résidu est négatif, le point est en dessous de la droite.

(c) La somme des résidus est nulle : $e_4 = -(0{,}4 - 1{,}2 + 0{,}5) = 0{,}3$.

(d) $1^2 + (-2)^2 + 0{,}5^2 + 0{,}5^2 = 1 + 4 + 0{,}25 + 0{,}25 = 5{,}5$.

(e) Faux : elle rend minimale la somme des carrés des résidus. La somme des résidus vaut 0 pour toute droite qui passe par G, ce critère ne choisit donc pas une seule droite.

(f) $0{,}8 \times 15 + 12 = 24$. On résout l'équation de la droite donnée : $0{,}8x + 12 = 20$, d'où $x = \frac{8}{0{,}8} = 10$.

Ex. 17 ★★☆ **Publicité et chiffre d'affaires**

(a) $\overline{x} = \frac{2+3+6+9}{4} = 5$ et $\overline{y} = \frac{31+35+40+46}{4} = 38$: $G(5\,;38)$.

(b) Écarts : $x_i - \overline{x}$: $-3\,;-2\,;1\,;4$ et $y_i - \overline{y}$: $-7\,;-3\,;2\,;8$.
$\text{cov}(x,y) = \frac{21+6+2+32}{4} = 15{,}25$ et $V(x) = \frac{9+4+1+16}{4} = 7{,}5$.

(c) $a = \frac{15{,}25}{7{,}5} = \frac{61}{30} \approx 2{,}033$ et $b = 38 - \frac{61}{30} \times 5 = \frac{167}{6} \approx 27{,}833$: $y = 2{,}033x + 27{,}833$. Contrôle : $\frac{61}{30} \times 5 + \frac{167}{6} = \frac{61+167}{6} = 38$, la droite passe par G.

(d) $x = 7$: $\frac{61}{30} \times 7 + \frac{167}{6} = \frac{1262}{30} \approx 42{,}067$. Selon le modèle, le chiffre d'affaires serait d'environ 42 070 €.

Ex. 17 (suite)

(e) On résout l'équation de la droite : $\frac{61}{30}x + \frac{167}{6} = 50$, soit $\frac{61}{30}x = \frac{133}{6}$, d'où $61x = 665$ et $x \approx 10{,}902$. Il faudrait un budget d'environ 1 090 €. C'est une extrapolation (on sort de $[2\,;9]$), l'estimation est donc peu fiable.

Ex. 18 ★★☆ L'erreur de Chloé

(a) $G(6\,;50)$ et $4 \times 6 + 74 = 98 \neq 50$: la droite ne passe pas par G, ce n'est donc pas la droite des moindres carrés.

(b) Elle divise par σ_x au lieu de $V(x) = \sigma_x^2$, et elle écrit $b = \overline{y} + a\overline{x}$ au lieu de $b = \overline{y} - a\overline{x}$.

(c) $V(x) = 2^2 = 4$, donc $a = \frac{8}{4} = 2$ et $b = 50 - 2 \times 6 = 38$: $y = 2x + 38$. Contrôle : $2 \times 6 + 38 = 50$, la droite passe bien par G.

(d) $2 \times 8 + 38 = 54$: selon le modèle, ce lycéen obtiendrait environ 54 points sur 100.

(e) Non : une droite des moindres carrés existe pour tout nuage. Les points ne sont alignés que si tous les résidus sont nuls ; ajuster, c'est seulement approcher le nuage par une droite.

Ex. 19 ★★☆ Deux droites en concurrence

(a) y : 12 ; 15 ; 15 ; 18, donc $G(2{,}5\,;15)$. $1{,}8 \times 2{,}5 + 10{,}5 = 15$ et $2 \times 2{,}5 + 10 = 15$: les deux droites passent par G.

(b) D_1 : valeurs prédites 12,3 ; 14,1 ; 15,9 ; 17,7, résidus $-0{,}3\,;0{,}9\,;-0{,}9\,;0{,}3$ et $S_1 = 0{,}09 + 0{,}81 + 0{,}81 + 0{,}09 = 1{,}8$. D_2 : résidus $0\,;1\,;-1\,;0$ et $S_2 = 2$.

(c) Non. On a bien $\sum |e_i| = 2{,}4$ pour D_1 et 2 pour D_2, mais la méthode des moindres carrés compare la somme des carrés des résidus : $S_1 = 1{,}8 < S_2 = 2$, donc c'est D_1 la meilleure en ce sens.

(d) Les $y_i - 1{,}8x_i$ valent 10,2 ; 11,4 ; 9,6 ; 10,8 (somme 42, somme des carrés 442,8).
Donc $S(b) = 4b^2 - 84b + 442{,}8$. Trinôme de coefficient $4 > 0$: minimal pour $b = \frac{84}{2\times 4} = 10{,}5$. On retrouve D_1.

Ex. 20 ★★★ Démonstration guidée : pourquoi cette droite ?

(a) $\sum y_i = 3 + 6 + 8 + 11 = 28$ et $\sum x_i = 2 + 4 + 5 + 9 = 20$, donc $\sum(y_i - ax_i) = 28 - 20a$. On a alors $S = 4b^2 - 2(28 - 20a)b + \sum(y_i - ax_i)^2$. C'est un trinôme en b dont le coefficient 4 est positif. Il est minimal en $b = \frac{2(28-20a)}{2\times 4} = 7 - 5a$. Or $\overline{x} = \frac{20}{4} = 5$ et $\overline{y} = \frac{28}{4} = 7$, donc $b = \overline{y} - a\overline{x}$. Ainsi $a \times 5 + b = 7$: la meilleure droite passe par le point moyen $G(5\,;7)$.

Ex. 20 (suite)

(b) $y_i - ax_i - b = (y_i - 7) - a(x_i - 5)$. Les écarts $x_i - \overline{x}$ valent $-3\,;-1\,;0\,;4$ et les écarts $y_i - \overline{y}$ valent $-4\,;-1\,;1\,;4$. On calcule $\sum(x_i - \overline{x})^2 = 9 + 1 + 0 + 16 = 26$, puis $\sum(x_i - \overline{x})(y_i - \overline{y}) = 12 + 1 + 0 + 16 = 29$ et $\sum(y_i - \overline{y})^2 = 16 + 1 + 1 + 16 = 34$.
Donc $S = 26a^2 - 58a + 34$. Ce trinôme est minimal en $a = \frac{58}{2\times 26} = \frac{29}{26}$. Alors $b = 7 - 5 \times \frac{29}{26} = \frac{37}{26}$. Contrôle : $\mathrm{cov}(x,y) = \frac{29}{4} = 7{,}25$ et $V(x) = \frac{26}{4} = 6{,}5$, d'où $a = \frac{7{,}25}{6{,}5} = \frac{29}{26}$. Au millième, la droite des moindres carrés est $y = 1{,}115x + 1{,}423$.

(c) $\sigma_x = \sqrt{6{,}5}$, donc Yanis obtient $a = \frac{7{,}25}{\sqrt{6{,}5}} \approx 2{,}84$. Avec cette valeur, $S = 26a^2 - 58a + 34 \approx 79{,}32$. Or le trinôme S atteint son minimum seulement en $a = \frac{29}{26}$, où $S = \frac{43}{26} \approx 1{,}65$. La valeur de Yanis ne minimise donc pas la somme des carrés des résidus. Le bon coefficient est $a = \frac{\mathrm{cov}(x,y)}{V(x)}$, pas $\frac{\mathrm{cov}(x,y)}{\sigma_x}$.

(d) Elle a tort. Les résidus de $y = x + 2$ valent $3 - 4 = -1$, $6 - 6 = 0$, $8 - 7 = 1$ et $11 - 11 = 0$. Leur somme est bien nulle, mais toute droite qui passe par G a une somme de résidus nulle, et c'est le cas ici ($5 + 2 = 7$). Le critère est la somme des CARRÉS des résidus. Ici elle vaut $1 + 0 + 1 + 0 = 2$, ce qui est plus que $\frac{43}{26} \approx 1{,}65$. Cette droite n'est donc pas la droite des moindres carrés.

(e) Ilan prend $a \approx 1{,}12$ et trouve $b = 7 - 5 \times 1{,}12 = 1{,}4$. La bonne valeur est $b = \frac{37}{26} \approx 1{,}423$. Même avec a arrondi au millième, on obtiendrait $7 - 5 \times 1{,}115 = 1{,}425$, qui reste faux. On calcule b avec la valeur exacte de a, et on n'arrondit qu'à la fin.

Ex. 21 ★★★ Vélos en libre-service : du modèle à la prévision

(a) $\overline{x} = \frac{1+2+3+4+5+6}{6} = 3{,}5$ et $\overline{y} = \frac{37{,}8}{6} = 6{,}3$, donc $G(3{,}5\,;6{,}3)$. La calculatrice (régression linéaire) donne $a = 0{,}857142\ldots$ et $b = 3{,}3$. La droite des moindres carrés est donc $y = 0{,}857x + 3{,}3$. Contrôle, avec a non arrondi : $0{,}857142\ldots \times 3{,}5 + 3{,}3 = 3 + 3{,}3 = 6{,}3 = \overline{y}$. La droite passe bien par G.

(b) Il a tort. Pour $x = 3$, avec a non arrondi, la valeur estimée est $0{,}857142\ldots \times 3 + 3{,}3 \approx 5{,}871$. Le résidu vaut $5{,}7 - 5{,}871 = -0{,}171$: il n'est pas nul. Comme il est négatif, le point est sous la droite et non dessus. Une valeur de r proche de 1 montre une forte corrélation linéaire positive, ce qui justifie l'ajustement affine. Mais ajuster ne veut pas dire que les points sont alignés.

Ex. 21 (suite)

(c) Pour $x = 6$, avec a non arrondi, la valeur estimée est $0{,}857142\ldots \times 6 + 3{,}3 \approx 8{,}443$. Le résidu vaut $8{,}6 - 8{,}443 \approx 0{,}157$: le point est au-dessus de la droite. La somme des six résidus vaut 0. En effet, la droite passe par G, donc $\sum e_i = \sum(y_i - \overline{y}) - a\sum(x_i - \overline{x}) = 0 - 0 = 0$.

(d) Le modèle retenu est la droite de y en x. Pour trouver x à partir d'une valeur de y, on résout l'équation (ici l'inéquation) de cette droite. La droite de x en y est une autre droite, qui ne correspond pas au modèle choisi. On résout $0{,}857142\ldots \times x + 3{,}3 \geqslant 12$, soit $x \geqslant \frac{12-3{,}3}{0{,}857142\ldots} = 10{,}15$. Le premier rang entier qui convient est $x = 11$, et $2018 + 11 = 2029$. Selon ce modèle, le service atteindrait 12 000 abonnés en 2029.

(e) Il faut être prudent. D'abord, c'est une extrapolation : le rang 11 est loin des données, qui vont du rang 1 au rang 6. Ensuite, rien ne garantit que la hausse restera affine aussi longtemps : le nombre d'abonnés peut plafonner, et les tarifs, le nombre de vélos ou la concurrence peuvent changer. La prévision 2029 est donc une simple tendance.

4. *Coefficient de corrélation, corrélation et causalité*

Ex. 22 ★☆☆ **Automatismes : vrai ou faux sur r**

(a) Faux : on a toujours $-1 \leqslant r \leqslant 1$. La valeur 1,08 vient d'une erreur de saisie ou de calcul.

(b) Vrai : si $r = 0$, alors $\text{cov}(x, y) = 0$, donc $a = \frac{\text{cov}(x,y)}{V(x)} = 0$. La droite a pour équation $y = \overline{y}$: elle est parallèle à l'axe des abscisses.

(c) Faux : r n'est pas un pourcentage. Il mesure seulement à quel point les points sont proches d'une droite. Avec $r = 0{,}95$, il est même possible qu'aucun point ne soit exactement sur la droite.

(d) Faux. On a $\overline{x} = 4$ et $\sum(x_i - 4) = 0$, donc $\text{cov}(x, y) = \frac{(-3)\times2+(-2)\times8+(-1)\times12+0\times14+1\times12+2\times8+3\times2}{7} = 0$ et $r = 0$. Mais le nuage montre un lien très net, en forme de cloche, avec un maximum en juin. Il n'y a pas de lien affine, mais il y a bien un lien.

(e) Faux : passer en mètres revient à multiplier x par $1\,000 > 0$, et ce changement d'unité ne modifie pas r.

(f) Vrai : $r < 0$ indique un nuage décroissant, et $|r| = 0{,}98$ est proche de 1, donc les points sont presque alignés.

Ex. 22 (suite)

(g) Faux : une corrélation ne prouve pas une causalité. Ici, une variable cachée, l'âge, fait augmenter à la fois la pointure et la vitesse de lecture.

Ex. 23 ★★☆ **Automatismes : calculer r et l'exploiter**

(a) $r = \frac{6{,}3}{2{,}5\times2{,}8} = \frac{6{,}3}{7} = 0{,}9$.

(b) $\sigma_x = \sqrt{16} = 4$ et $\sigma_y = \sqrt{16{,}2}$, donc $r = \frac{-14{,}4}{4\sqrt{16{,}2}} \approx -0{,}894$.

(c) $r = \sqrt{0{,}8836}$ ou $r = -\sqrt{0{,}8836}$. Comme r a le signe de $a = -2{,}1 < 0$, on a $r = -\sqrt{0{,}8836} = -0{,}94$.

(d) $\text{cov}(x, y) = r\,\sigma_x\sigma_y = 0{,}8 \times 2 \times 5 = 8$ et $V(x) = 2^2 = 4$, donc $a = \frac{\text{cov}(x,y)}{V(x)} = \frac{8}{4} = 2$.

(e) Léna a arrondi trop tôt. On garde les valeurs exactes : $r = \frac{5{,}4}{\sqrt{3}\times\sqrt{10}} = \frac{5{,}4}{\sqrt{30}} \approx 0{,}986$, et non 0,993.

(f) Non : $a = \frac{\text{cov}(x,y)}{V(x)}$ avec $V(x) > 0$, donc a a le signe de la covariance. Si $\text{cov}(x, y) = -3$, alors $a < 0$, ce qui exclut $a = 0{,}5$.

Ex. 24 ★★☆ **Entraînement et temps au 3 000 m**

(a) $\overline{x} = \frac{2+4+6+8}{4} = 5$ et $\overline{y} = \frac{15{,}4+14{,}6+13{,}9+12{,}9}{4} = 14{,}2$, donc $G(5; 14{,}2)$. Ensuite $\text{cov}(x, y) = \frac{(-3)\times1{,}2+(-1)\times0{,}4+1\times(-0{,}3)+3\times(-1{,}3)}{4} = \frac{-8{,}2}{4} = -2{,}05$.

(b) $V(x) = \frac{9+1+1+9}{4} = 5$ et $V(y) = \frac{1{,}44+0{,}16+0{,}09+1{,}69}{4} = 0{,}845$. Alors $r = \frac{-2{,}05}{\sqrt{5\times0{,}845}} \approx -0{,}997$.

(c) $r < 0$: le nuage est décroissant ; les coureurs qui s'entraînent le plus ont les temps les plus courts. Comme $|r| \approx 0{,}997 \geqslant 0{,}9$, un ajustement affine est justifié selon le critère du club.

(d) $a = \frac{\text{cov}(x,y)}{V(x)} = \frac{-2{,}05}{5} = -0{,}41$ et $b = \overline{y} - a\overline{x} = 14{,}2 + 0{,}41 \times 5 = 16{,}25$, donc $y = -0{,}41x + 16{,}25$. Pour $x = 20$: $y = -0{,}41 \times 20 + 16{,}25 = 8{,}05$ min. Or 20 est loin hors de $[2; 8]$: c'est une extrapolation, peu fiable malgré $|r| \approx 0{,}997$. Le temps ne peut pas baisser indéfiniment, puisque la droite donnerait même un temps négatif au-delà d'environ 40 h. On ne peut pas annoncer environ 8 min pour ce coureur.

Ex. 25 ★★☆ **Chauffage et point aberrant**

(a) À la calculatrice, en mode statistique : $r \approx -0{,}818$ et $y = -3{,}53x + 59{,}81$.

(b) $|r| \approx 0{,}818 < 0{,}9$, donc selon le seuil fixé, l'ajustement affine n'est pas justifié avec ces six relevés.

Ex. 25 **(suite)**

(c) Sans le jour 4 : $r \approx -0{,}995$ et $y = -3{,}20x + 61{,}98$. Un seul point aberrant faisait nettement chuter $|r|$. Maintenant $|r| \geqslant 0{,}9$, donc l'ajustement affine est justifié.

(d) Avec la première droite : $-3{,}53 \times 5 + 59{,}81 \approx 42{,}2$ kWh. Avec la seconde : $-3{,}20 \times 5 + 61{,}98 \approx 46{,}0$ kWh. Le point aberrant modifie aussi la droite et fausse l'estimation d'environ 4 kWh. On retient la seconde : il faut prévoir environ 46 kWh à 5 °C.

Ex. 26 ★★☆ **Piscines et cambriolages**

(a) À la calculatrice : $r \approx 0{,}991$. Comme $r > 0$ et $|r| \geqslant 0{,}9$, il y a une forte corrélation linéaire positive : les communes qui ont plus de piscines ont aussi plus de cambriolages.

(b) Non : une forte corrélation ne prouve pas une causalité. Il existe une variable cachée, la taille de la commune : une commune plus peuplée a plus de piscines et aussi plus de cambriolages.

(c) A : $\frac{40}{8} = 5$; B : $\frac{72}{15} = 4{,}8$; C : $\frac{112}{22} \approx 5{,}09$; D : $\frac{196}{40} = 4{,}9$; E : $\frac{330}{65} \approx 5{,}08$. Le taux est presque le même partout, environ 5 cambriolages pour 1 000 habitants. La population suffit à expliquer les variations de y.

(d) Les deux variables varient ensemble ($r \approx 0{,}991$), mais ces données ne permettent pas de dire que les piscines provoquent des cambriolages. Le lien s'explique par la taille des communes. Pour conclure à une causalité, il faudrait une expérience contrôlée ou un mécanisme connu.

Ex. 27 ★★★ **Le câble de la passerelle**

(a) La calculatrice donne $r \approx -0{,}011$. Ici $\overline{x} = 0$, $\sigma_x = 20$, $\overline{y} = 10$ et $\operatorname{cov}(x, y) = -\frac{11}{7} \approx -1{,}571$. Comme $|r| < 0{,}99$, il n'y a pas de lien affine. Pourtant, le tableau montre que y décroît puis croît de façon presque symétrique : le lien est très net, mais courbé. Mehdi a donc tort : r proche de 0 veut dire « pas de lien affine », pas « aucun lien ».

(b) Les z_i valent 900 ; 400 ; 100 ; 0 ; 100 ; 400 ; 900. La calculatrice donne $r \approx 0{,}9997$ entre z et y. Comme $|r| \geqslant 0{,}99$, un ajustement affine de y en z est pertinent.

(c) Par la méthode des moindres carrés, la calculatrice donne $y = 0{,}0201z + 1{,}976$. Comme $z = x^2$, on obtient $y \approx 0{,}0201x^2 + 1{,}976$: le câble a une forme parabolique.

Ex. 27 **(suite)**

(d) Pour $x = 25$, on a $z = 625$ et $y \approx 0{,}0200595 \times 625 + 1{,}97619 \approx 14{,}5$. La valeur 25 est dans l'intervalle $[-30; 30]$ des mesures : c'est une interpolation, donc l'estimation est fiable. À 25 m du milieu, le câble est à environ 14,5 m au-dessus du tablier.

(e) Changer d'unité revient à multiplier les données par un nombre positif : $y' = 100y$, $x' = 10x$ et donc $z' = x'^2 = 100z$. Le coefficient r ne change pas quand on multiplie par un nombre positif. Nour obtient donc encore $r \approx -0{,}011$ entre positions et hauteurs, et $r \approx 0{,}9997$ entre z et les hauteurs.

Ex. 28 ★★★ **Glaces et sauvetages en mer**

(a) On a $\overline{x} = 28$ et $\overline{y} = 7$. La calculatrice donne $r \approx 0{,}980$ et $y = 0{,}182x + 1{,}904$. Comme $r > 0$ et $|r| \geqslant 0{,}9$, la corrélation est forte et positive : les semaines où l'on vend plus de glaces sont aussi celles où il y a plus de sauvetages.

(b) Pour $x = 26$, la droite donne $0{,}182 \times 26 + 1{,}904 \approx 6{,}636$. Le résidu vaut $7 - 6{,}636 \approx 0{,}364$: le point n'est pas sur la droite. En fait, aucun des six points n'est sur la droite. Inès a tort : r mesure à quel point le nuage est aligné dans son ensemble, ce n'est pas un pourcentage de points sur la droite.

(c) La calculatrice donne $r \approx 0{,}994$ entre t et x, et $r \approx 0{,}976$ entre t et y. Les deux corrélations sont fortes et positives : les ventes de glaces et les sauvetages augmentent tous les deux avec la température.

(d) Non. Les valeurs 0,980 et 0,976 sont très proches et ne portent que sur six semaines. Surtout, r mesure seulement à quel point un nuage est proche d'une droite : un r plus grand ne désigne pas une cause. Acheter une glace ne provoque pas de sauvetage, alors que la chaleur pousse à la fois à acheter des glaces et à se baigner. Yanis ne peut pas conclure ainsi.

(e) La température est très fortement corrélée aux ventes de glaces ($r \approx 0{,}994$) et aux sauvetages ($r \approx 0{,}976$). C'est une variable cachée plausible : quand il fait chaud, on achète plus de glaces et on se baigne plus, d'où plus de sauvetages. La corrélation $r \approx 0{,}980$ entre x et y ne prouve donc pas que les glaces causent les sauvetages. En plus, $n = 6$ est petit. La mesure proposée par l'élu n'est pas justifiée par ces données : seule une expérience contrôlée pourrait établir un lien de cause à effet.

5. *Interpoler et extrapoler avec un ajustement*

Ex. 29 ★☆☆ **Automatismes : interpolation ou extrapolation ?**

(a) $y = -1{,}8 \times 3 + 52 = 46{,}6$, soit environ 46 600 abonnés. Comme $3 \in [1\,;7]$, c'est une interpolation.

(b) $x = 2027 - 2018 + 1 = 10$ et $y = -1{,}8 \times 10 + 52 = 34$, soit environ 34 000 abonnés en 2027. Comme $10 \notin [1\,;7]$, c'est une extrapolation, valable seulement si la tendance se poursuit.

(c) 2017 correspond à $x = 0$ et $y = -1{,}8 \times 0 + 52 = 52$, soit environ 52 000 abonnés. Comme $0 \notin [1\,;7]$, c'est une extrapolation, mais elle reste proche de la plage.

(d) $x = 2047 - 2018 + 1 = 30$ et $y = -1{,}8 \times 30 + 52 = -2$. C'est une extrapolation lointaine et le résultat est absurde, puisqu'un nombre d'abonnés ne peut pas être négatif. L'ajustement n'a plus de sens aussi loin de la plage.

Ex. 30 ★☆☆ **Automatismes : atteindre un seuil**

(a) $2{,}5x > 18 \iff x > \frac{18}{2{,}5} = 7{,}2$. Le plus petit entier qui convient est $x = 8$.

(b) $-0{,}6x < -5 \iff x > \frac{-5}{-0{,}6} \approx 8{,}33$: on divise par un nombre négatif, donc le sens change. Le plus petit entier qui convient est $x = 9$.

(c) $1{,}35x > 9{,}8 \iff x > \frac{9{,}8}{1{,}35} \approx 7{,}26$, donc $x = 8$. L'année est $2021 + 8 - 1 = 2028$: y dépasse 50 pour la première fois en 2028.

(d) Faux : pour $x = 10$, $y = -3{,}2 \times 10 + 75 = 43 > 40$. On résout $-3{,}2x + 75 \leqslant 40 \iff -3{,}2x \leqslant -35 \iff x \geqslant \frac{-35}{-3{,}2} = 10{,}9375$ (on divise par un nombre négatif, donc le sens change). Le plus petit entier qui convient est $x = 11$.

(e) $-0{,}4x > -6$. Noé a divisé par $-0{,}4 < 0$ sans changer le sens de l'inéquation. La bonne réponse est $x < \frac{-6}{-0{,}4} = 15$, donc $x < 15$.

Ex. 31 ★★☆ **Une coopérative solaire à Aurillac**

(a) Par la méthode des moindres carrés, la calculatrice donne $y \approx 14{,}57x + 105{,}67$.

(b) 2023 correspond à $x = 2023 - 2015 + 1 = 9$. Avec les valeurs non arrondies, $y \approx 14{,}571 \times 9 + 105{,}667 \approx 236{,}81$. La production estimée est d'environ 236,81 MWh. Comme $9 \notin [1\,;6]$, c'est une extrapolation, valable seulement si la tendance se poursuit.

Ex. 31 (suite)

(c) $\frac{236{,}81 - 228}{228} \times 100 \approx 3{,}86$. La prévision dépasse la valeur réelle d'environ 3,86 % : une estimation n'est pas une valeur certaine.

(d) $14{,}571x + 105{,}667 > 250 \iff x > \frac{250 - 105{,}667}{14{,}571} \approx 9{,}91$, donc $x = 10$. Si la tendance se poursuit, la production dépasse 250 MWh à partir de 2024.

Ex. 32 ★★☆ **La fibre à Millau**

(a) $\overline{x} = \frac{1+2+3+4+5}{5} = 3$ et $\overline{y} = \frac{62+68+73+79+85}{5} = 73{,}4$, donc $G(3\,;73{,}4)$.

(b) La droite passe par G, donc $b = \overline{y} - a\overline{x} = 73{,}4 - 5{,}7 \times 3 = 56{,}3$. L'ajustement est $y = 5{,}7x + 56{,}3$.

(c) 2029 correspond à $x = 2029 - 2020 + 1 = 10$ et $y = 5{,}7 \times 10 + 56{,}3 = 113{,}3$. Un taux de 113,3 % est absurde, puisqu'un taux ne peut pas dépasser 100 %. Cette extrapolation va trop loin.

(d) $5{,}7x + 56{,}3 > 100 \iff x > \frac{43{,}7}{5{,}7} \approx 7{,}67$, donc $x = 8$, c'est-à-dire 2027. À partir de 2027, le modèle affine n'a plus de sens : le taux va forcément plafonner.

Ex. 33 ★★☆ **La chaîne vidéo de Maëlys**

(a) $\ln y = 0{,}18x + 2{,}1$, donc $y = \mathrm{e}^{0{,}18x + 2{,}1} = \mathrm{e}^{2{,}1} \times \mathrm{e}^{0{,}18x}$, avec $k = \mathrm{e}^{2{,}1} \approx 8{,}17$. On obtient $y \approx 8{,}17\,\mathrm{e}^{0{,}18x}$.

(b) $y = \mathrm{e}^{0{,}18 \times 6 + 2{,}1} = \mathrm{e}^{3{,}18} \approx 24{,}05$, soit environ 24 050 abonnés au mois 6. Comme $6 \in [1\,;8]$, c'est une interpolation.

(c) $y > 100 \iff 0{,}18x + 2{,}1 > \ln 100 \iff x > \frac{\ln 100 - 2{,}1}{0{,}18} \approx 13{,}92$, donc $x = 14$. Si la tendance se poursuit, la chaîne dépasse 100 000 abonnés à partir du mois 14 (c'est une extrapolation).

(d) Non. Le calcul $\mathrm{e}^{0{,}18 \times 24 + 2{,}1} = \mathrm{e}^{6{,}42} \approx 614$ est juste. Mais le mois 24 est très loin de la plage $[1\,;8]$: c'est une extrapolation lointaine, donc une estimation fragile et pas une valeur certaine. On peut seulement dire : « environ 614 000 abonnés, si la tendance se poursuit ».

Ex. 34 ★★★ **Enneigement près de Font-Romeu : jusqu'où croire le modèle ?**

(a) À la calculatrice : $y = -5{,}21x + 123{,}14$ et $r \approx -0{,}996$. Comme r est très proche de -1, la corrélation linéaire est forte et négative : un ajustement affine est justifié.

(b) L'hiver 2027 correspond à $x = 2027 - 2016 = 11$. $y = -5{,}21 \times 11 + 123{,}14 = 65{,}83$. Comme $11 \notin [1\,;7]$, c'est une extrapolation : si la tendance se poursuit, l'hiver 2027 compterait environ 66 jours d'enneigement.

Ex. 34 **(suite)**

(c) Pour 2025, $x = 9$ et $y = -5{,}21 \times 9 + 123{,}14 = 76{,}25$, soit environ 76 jours. Écart relatif : $\frac{76{,}25-71}{71} \approx 0{,}074$, soit environ 7,4 %. Le modèle surestime l'enneigement de 2025 : une estimation n'est pas une valeur certaine.

(d) $-5{,}21x + 123{,}14 < 50 \iff -5{,}21x < -73{,}14 \iff x > \frac{73{,}14}{5{,}21} \approx 14{,}04$ (on divise par $-5{,}21 < 0$: le sens change). Le plus petit entier qui convient est $x = 15$ (pour $x = 14$, $y = 50{,}20$; pour $x = 15$, $y = 44{,}99$). Le rang 15 correspond à $2016 + 15 = 2031$: si la tendance se poursuit, l'enneigement passerait sous 50 jours à partir de l'hiver 2031.

(e) Faux. 2039 correspond à $x = 23$ et $y = -5{,}21 \times 23 + 123{,}14 = 3{,}31 > 0$. On a $y < 0 \iff x > \frac{123{,}14}{5{,}21} \approx 23{,}64$, donc le modèle donne une valeur négative à partir de $x = 24$, soit l'hiver 2040 ($y = -1{,}90$). Or un nombre de jours ne peut pas être négatif : une extrapolation aussi éloignée de la plage $[1\,;7]$ n'a plus de sens.

Ex. 35 ★★★ **Une culture de levures : affine ou exponentiel ?**

(a) À la calculatrice : $y = 12x - 4{,}4$. Pour $x = 3{,}5$: $y = 12 \times 3{,}5 - 4{,}4 = 37{,}6$. Comme $3{,}5 \in [1\,;5]$, c'est une interpolation : à 3 h 30, la concentration est estimée à environ 37 600 cellules par mL.

(b) On calcule $z_i = \ln y_i$, puis à la calculatrice : $z = 0{,}405x + 2{,}079$. Alors $y = \mathrm{e}^{z} = \mathrm{e}^{2{,}079} \times \mathrm{e}^{0{,}405x}$ avec $\mathrm{e}^{2{,}079} \approx 8$, donc $y = 8\,\mathrm{e}^{0{,}405x}$.

(c) Modèle affine : $12 \times 8 - 4{,}4 = 91{,}6$, écart $\frac{91{,}6-198}{198} \approx -53{,}7\,\%$. Modèle exponentiel : $8\,\mathrm{e}^{0{,}405\times 8} \approx 204{,}27$, écart $\frac{204{,}27-198}{198} \approx 3{,}2\,\%$. Hors de la plage $[1\,;5]$, le modèle exponentiel prévoit bien mieux : le modèle affine sous-estime fortement la croissance.

(d) $8\,\mathrm{e}^{0{,}405x} > 500 \iff \mathrm{e}^{0{,}405x} > 62{,}5 \iff x > \frac{\ln 62{,}5}{0{,}405} \approx 10{,}21$. Le plus petit entier qui convient est $x = 11$ (pour $x = 10$, $y \approx 459{,}2$; pour $x = 11$, $y \approx 688{,}4$). Si la tendance se poursuit, la concentration dépasse 500 milliers de cellules par mL à partir de 11 h.

(e) Faux. Le modèle donne $8\,\mathrm{e}^{0{,}405\times 24} \approx 133\,000$ milliers, soit environ 133 millions de cellules par mL : c'est seulement une estimation. Comme 24 est très loin de la plage $[1\,;5]$, c'est une extrapolation lointaine qui suppose que la croissance se poursuit au même rythme. Or une culture finit par manquer de nutriments, donc rien n'est « certain ».

6. *Ajustement par changement de variable*

Ex. 36 ★☆☆ **Automatismes : passer par $z = \ln y$ et revenir à y**

(a) Non : $\ln y$ n'existe que si $y > 0$, et $\ln 0$ n'existe pas. On ne peut pas poser $z = \ln y$ pour cette série.

(b) $z = \ln 45 \approx 3{,}807$.

(c) $y = \mathrm{e}^{0{,}2x+1{,}5} = \mathrm{e}^{1{,}5} \times (\mathrm{e}^{0{,}2})^{x} \approx 4{,}48 \times 1{,}22^{x}$.

(d) Non. Pour $x = 2$, le modèle donne $y = \mathrm{e}^{1{,}9} \approx 6{,}69$, alors que Léa obtient $\mathrm{e}^{0{,}4} + \mathrm{e}^{1{,}5} \approx 5{,}97$. En effet, $\mathrm{e}^{0{,}2x+1{,}5} = \mathrm{e}^{1{,}5} \times \mathrm{e}^{0{,}2x}$: c'est un produit, pas une somme.

(e) 3,1 est une estimation de $z = \ln y$, pas de y. On revient à y : $y = \mathrm{e}^{3{,}1} \approx 22{,}20$.

(f) Quand x augmente de 1, y est multiplié par $\mathrm{e}^{-0{,}05} \approx 0{,}9512$. Comme $\mathrm{e}^{-0{,}05} - 1 \approx -0{,}0488$, y baisse d'environ 4,88 %.

Ex. 37 ★☆☆ **Automatismes : choisir le bon changement de variable**

(a) Une hyperbole suggère un modèle $y = \frac{\alpha}{x} + \beta$: on pose $z = \frac{1}{x}$ et on ajuste y en z.

(b) Ici $z = \frac{1}{x}$. Pour $x = 8$, on a $z = \frac{1}{8}$, donc $y = \frac{24}{8} + 1{,}5 = 4{,}5$.

(c) $\ln y = \ln k + p \ln x$: on pose $u = \ln x$ et $v = \ln y$. On ajuste alors $v = pu + \ln k$.

(d) Ici $X = \ln x$ et $Y = \ln y$, donc $Y = pX + \ln k$. Ainsi $p = 1{,}5$ et $\ln k = 0{,}7$, donc $k = \mathrm{e}^{0{,}7} \approx 2{,}01$. Le modèle s'écrit $y \approx 2{,}01 \times x^{1{,}5}$.

(e) $y = \mathrm{e}^{0{,}7} \times 4^{1{,}5} = \mathrm{e}^{0{,}7} \times 8 \approx 16{,}11$.

(f) Faux. Un coefficient multiplicateur constant correspond à un modèle exponentiel : on pose $z = \ln y$. Le coefficient directeur attendu vaut environ $\ln 1{,}08 \approx 0{,}077$.

(g) $\mathrm{e}^{0{,}3x+2} \geqslant 100 \iff 0{,}3x + 2 \geqslant \ln 100$, car ln est croissante.
Donc $x \geqslant \frac{\ln 100 - 2}{0{,}3} \approx 8{,}68$.

Ex. 38 ★★☆ **Élimination d'un médicament**

(a) Toutes les valeurs y_i sont strictement positives, donc $z = \ln y$ existe. $z \approx 6{,}174$; 5,991 ; 5,753 ; 5,529 ; 5,298 ; 5,088. La calculatrice donne $z = -0{,}221x + 6{,}191$, avec $r \approx -1{,}00$. L'ajustement affine de z en x est donc très pertinent.

(b) $y = \mathrm{e}^{-0{,}221x+6{,}191} = \mathrm{e}^{6{,}191} \times (\mathrm{e}^{-0{,}221})^{x} \approx 488 \times 0{,}802^{x}$.

Ex. 38 **(suite)**

(c) $\mathrm{e}^{-0{,}221} - 1 \approx -0{,}198$. Selon ce modèle, la concentration baisse d'environ 19,8 % par heure.

(d) $\mathrm{e}^{-0{,}221x+6{,}191} < 50 \iff -0{,}221x + 6{,}191 < \ln 50 \iff x > \frac{6{,}191 - \ln 50}{0{,}221} \approx 10{,}3$. La concentration passe sous 50 μg/L au bout d'environ 10,3 h. C'est une extrapolation (les mesures s'arrêtent à 5 h) : le résultat reste à confirmer.

Ex. 39 ★★☆ **Panneaux solaires : affine ou exponentiel ?**

(a) $z \approx 2{,}996$; 3,258 ; 3,555 ; 3,784 ; 4,060 ; 4,317. La calculatrice donne $z = 0{,}264x + 2{,}737$, avec $r \approx 1{,}000$: le nuage $(x\,;z)$ est presque aligné.

(b) $y = \mathrm{e}^{2{,}737} \times (\mathrm{e}^{0{,}264})^x \approx 15{,}4 \times 1{,}302^x$. Comme $1{,}302 - 1 = 0{,}302$, le modèle prévoit une hausse d'environ 30,2 % par an.

(c) Ajustement affine : $10{,}86 \times 7 + 5 \approx 81$ kWc, soit un écart de 15 kWc. Modèle exponentiel : $\mathrm{e}^{0{,}264\times 7+2{,}737} \approx 98$ kWc, soit un écart d'environ 2 kWc. L'ajustement exponentiel est le plus fiable.

(d) Pour $x = 7$, Noé obtient $z = 4{,}8$ au lieu de $0{,}264 \times 7 + 2{,}737 \approx 4{,}585$. L'écart d'environ 0,215 sur z multiplie y par $\mathrm{e}^{0{,}215} \approx 1{,}24$. L'exponentielle amplifie les erreurs : il faut garder au moins trois décimales pour a et b. On trouve alors environ 98 kWc.

Ex. 40 ★★☆ **Recharge d'une voiture électrique**

(a) $z_i = 0{,}1$; 0,05 ; 0,025 ; 0,02. $\overline{z} = \frac{0{,}195}{4} = 0{,}04875$ et $\overline{y} = \frac{516}{4} = 129$, donc $G(0{,}04875\,;129)$.

(b) $V(z) = \frac{0{,}013525}{4} - 0{,}04875^2 \approx 0{,}0010047$ et $\mathrm{cov}(z,y) = \frac{34{,}72}{4} - 0{,}04875 \times 129 = 2{,}39125$. $\alpha = \frac{2{,}39125}{0{,}0010047} \approx 2380$ et $\beta = 129 - \alpha \times 0{,}04875 \approx 13{,}0$. La droite est $y = 2380z + 13$.

(c) $y = \frac{2380}{x} + 13$. Pour $x = 25$: $y \approx \frac{2380}{25} + 13 \approx 108$. Sur une borne de 25 kW, la recharge dure environ 108 min (c'est une interpolation, donc fiable).

(d) $\frac{2380}{x} + 13 < 45 \iff \frac{2380}{x} < 32 \iff x > \frac{2380}{32} \approx 74{,}4$, car $x > 0$. Il faut une borne de plus de 74,4 kW environ (en pratique 75 kW). C'est une extrapolation (au-delà de 50 kW) : le résultat est peu fiable.

Ex. 41 ★★★ **Croissance des bars : un modèle puissance**

(a) $\ln y = \ln(k\,x^p) = \ln k + \ln(x^p) = p \ln x + \ln k$. Avec $u = \ln x$ et $v = \ln y$, on a $v = p\,u + \ln k$: v est une fonction affine de u. Les x_i (longueurs) et les y_i (masses) sont strictement positifs, donc leurs logarithmes existent.

(b) $u_i \approx 2{,}996$; 3,219 ; 3,401 ; 3,555 ; 3,689 et $v_i \approx 4{,}585$; 5,220 ; 5,799 ; 6,234 ; 6,653. La calculatrice donne $v = 2{,}987u - 4{,}374$, avec $r \approx 0{,}9998$. Comme r est très proche de 1, l'ajustement affine de v en u est justifié, donc le modèle puissance aussi.

(c) $y = \mathrm{e}^v = \mathrm{e}^{2{,}987 \ln x - 4{,}374} = \mathrm{e}^{-4{,}374} \times \mathrm{e}^{2{,}987 \ln x} = \mathrm{e}^{-4{,}374} \times x^{2{,}987}$. L'exponentielle d'une somme est un PRODUIT d'exponentielles, pas une somme : Malo s'est trompé sur ce point. $k = \mathrm{e}^{-4{,}374} \approx 0{,}0126$, donc $y \approx 0{,}0126\,x^{2{,}987}$.

(d) $\ln y \approx -4{,}374 + 2{,}987 \times \ln 50 \approx 7{,}311$, donc $y \approx \mathrm{e}^{7{,}311} \approx 1\,497$ g. Ninon, elle, calcule $\mathrm{e}^{-4{,}4} \times 50^3 \approx 1\,535$ g. Une petite erreur sur p est multipliée par $\ln 50 \approx 3{,}9$, puis amplifiée par l'exponentielle : il faut garder assez de décimales. De plus, 50 cm est hors de $[20\,;40]$: c'est une extrapolation, à prendre avec prudence. Un bar de 50 cm pèse donc environ 1 497 g selon le modèle.

(e) $y \geqslant 1\,000 \iff \ln k + p \ln x \geqslant \ln 1\,000 \iff \ln x \geqslant \frac{\ln 1000 + 4{,}374}{2{,}987} \approx 3{,}777 \iff x \geqslant \mathrm{e}^{3{,}777} \approx 43{,}68$ (on peut diviser par $2{,}987 > 0$, et exp est croissante). Un bar dépasse donc 1 kg à partir d'environ 43,7 cm. C'est une légère extrapolation (au-delà de 40 cm), mais elle reste proche des données : l'estimation est assez fiable.

Ex. 42 ★★★ **Cote d'une citadine d'occasion**

(a) La calculatrice donne $y = -2{,}25x + 21{,}99$ et $r \approx -0{,}993$. $|r|$ est proche de 1, mais le nuage est courbé : les baisses successives diminuent (3,1 ; 2,5 ; 2,3 ; 1,8 ; 1,6). Un r proche de ± 1 ne suffit donc pas à valider l'ajustement affine : la forme du nuage suggère plutôt un modèle exponentiel.

(b) $z_i \approx 3{,}016$; 2,851 ; 2,695 ; 2,526 ; 2,370 ; 2,208. Toutes les cotes sont strictement positives, donc $\ln y$ existe. La calculatrice donne $z = -0{,}1613x + 3{,}1755$, avec $r \approx -1{,}000$: les points $(x_i\,;z_i)$ sont presque alignés, mieux encore que pour y.

(c) $y = \mathrm{e}^z = \mathrm{e}^{-0{,}1613x+3{,}1755} = \mathrm{e}^{3{,}1755} \times \left(\mathrm{e}^{-0{,}1613}\right)^x \approx 23{,}94 \times 0{,}851^x$. Le taux annuel vaut $\mathrm{e}^{-0{,}1613} - 1 \approx -0{,}149$: selon ce modèle, la cote baisse d'environ 14,9 % par an.

Ex. 42 **(suite)**

(d) Exponentiel : $\mathrm{e}^{-0{,}1613\times 8+3{,}1755} \approx 6{,}59$ k€, soit un écart de $0{,}01$ k€ avec la cote observée. Affine : $-2{,}25\times 8+21{,}99 = 3{,}99$ k€, soit un écart de $2{,}61$ k€. Le modèle exponentiel est bien plus fiable. Pour 10 ans, le modèle affine donne $-2{,}25\times 10+21{,}99 = -0{,}51$ k€, une cote négative : c'est absurde, et cela montre que ce modèle ne peut pas servir à extrapoler.

(e) $y<5 \iff -0{,}1613x+3{,}1755<\ln 5 \iff x>\frac{\ln 5-3{,}1755}{-0{,}1613}\approx 9{,}71$ (on divise par un nombre négatif, donc le sens de l'inégalité change). Vérification : pour $x=9$, $y\approx 5{,}61$ k€ ; pour $x=10$, $y\approx 4{,}77$ k€. Selon le modèle, la cote passe sous $5\,000$ € à partir de 10 ans.