1re tronc commun / Statistiques à deux variables — corrigé

benprof particulier.fr

1. *Automatismes : moyenne, médiane, quartiles*

Ex. 1 ★☆☆ Temps de trajet jusqu'au lycée

(a) $\overline{x} = \frac{18+7+25+12+9+30+15+12}{8} = \frac{128}{8} = 16$ minutes.

(b) Série rangée : 7 ; 9 ; 12 ; 12 ; 15 ; 18 ; 25 ; 30.

(c) $N = 8$ est pair : on fait la moyenne des valeurs de rangs 4 et 5, donc $Me = \frac{12+15}{2} = 13{,}5$ minutes.

(d) $\frac{8}{4} = 2$, on prend la valeur de rang 2 : $Q_1 = 9$ minutes.

(e) $\frac{3\times 8}{4} = 6$, on prend la valeur de rang 6 : $Q_3 = 18$ minutes.

(f) Écart interquartile : $Q_3 - Q_1 = 18 - 9 = 9$ minutes ; étendue : $30 - 7 = 23$ minutes.

Ex. 2 ★☆☆ Emprunts à la médiathèque

(a) $N = 7 + 6 + 8 + 4 = 25$ adhérents.

(b) $\overline{x} = \frac{1\times 7+2\times 6+3\times 8+4\times 4}{25} = \frac{59}{25} = 2{,}36$ livres par adhérent (et non $\frac{1+2+3+4}{4} = 2{,}5$: on pondère par les effectifs).

(c) Effectifs cumulés croissants : 7 ; 13 ; 21 ; 25.

(d) $N = 25$ est impair : la médiane est la valeur de rang 13. Les rangs 8 à 13 valent 2, donc $Me = 2$ livres.

(e) $\frac{25}{4} = 6{,}25$, on prend le rang 7 : les rangs 1 à 7 valent 1, donc $Q_1 = 1$ livre.

(f) $\frac{3\times 25}{4} = 18{,}75$, on prend le rang 19 : les rangs 14 à 21 valent 3, donc $Q_3 = 3$ livres. Écart interquartile : $3 - 1 = 2$ livres.

(g) $\frac{13}{25} = 0{,}52$, soit 52 % des adhérents.

Ex. 3 ★★☆ Les erreurs de Malo

(a) Malo n'a pas rangé la série. Série rangée : 0 ; 1 ; 2 ; 2 ; 3 ; 4 ; 5 ; 7. $N = 8$ est pair, donc $Me = \frac{2+3}{2} = 2{,}5$ buts.

(b) Non. $\frac{8}{4} = 2$: Q_1 est la valeur de rang 2 de la série rangée, on ne fait pas de moyenne.
Donc $Q_1 = 1$ but, et non 1,5.

(c) $\frac{3\times 8}{4} = 6$: Q_3 est la valeur de rang 6, donc $Q_3 = 4$ buts. Écart interquartile : $4 - 1 = 3$ buts.

(d) $\overline{x} = \frac{3+1+7+2+0+5+2+4}{8} = \frac{24}{8} = 3$ buts par match. La moyenne est un peu plus grande que la médiane (2,5) : le match à 7 buts la tire vers le haut.

Ex. 4 ★★☆ Salaire moyen ou salaire médian ?

(a) $\overline{x} = \frac{1650+1800+1720+2100+1900+1750+6400}{7} = \frac{17\,320}{7} \approx 2\,474$ €. Le salaire moyen est d'environ 2 474 €.

(b) Série rangée : 1 650 ; 1 720 ; 1 750 ; 1 800 ; 1 900 ; 2 100 ; 6 400. $N = 7$ est impair, la médiane est la 4e valeur : $Me = 1\,800$ €.

(c) Oui, c'est trompeur : la moyenne est tirée vers le haut par le seul salaire de 6 400 €. Six salariés sur sept gagnent moins de 2 500 € ; la médiane, 1 800 €, est plus représentative.

(d) La somme augmente de $8\,500 - 6\,400 = 2\,100$ €, donc la moyenne augmente de $\frac{2100}{7} = 300$ € : elle passe à environ 2 774 €. La plus grande valeur reste la 7e et la 4e ne change pas : la médiane reste 1 800 €.

Ex. 5 ★★☆ Un graphique qui exagère

(a) Parties visibles : $61\,200 - 60\,000 = 1\,200$ entrées, $62\,400 - 60\,000 = 2\,400$ entrées et $64\,800 - 60\,000 = 4\,800$ entrées.

(b) $\frac{4800}{1200} = 4$: la barre de 2025 paraît 4 fois plus haute que celle de 2023.

(c) Faux : $\frac{64800}{61200} \approx 1{,}06$. La fréquentation n'a été multipliée que par 1,06 environ ; l'axe tronqué fait paraître l'écart beaucoup plus grand.

(d) $\frac{64800-61200}{61200} \times 100 \approx 5{,}9$ %. Les entrées ont augmenté d'environ 5,9 % en deux ans. Il faut faire partir l'axe vertical de 0 : les barres seront alors presque de même hauteur.

Ex. 6 ★★★ Les trajets en bus d'une classe

(a) Effectifs cumulés croissants : 10 ; $10 + 9 = 19$; $19 + 7 = 26$; $26 + 4 = 30$. On retrouve bien l'effectif total $N = 30$.

(b) Inès fait la moyenne des valeurs distinctes sans tenir compte des effectifs : 2 trajets compte 10 fois, 8 trajets seulement 4 fois. Il faut une moyenne pondérée : $\overline{x} = \frac{10\times 2+9\times 4+7\times 6+4\times 8}{30} = \frac{130}{30} \approx 4{,}33$. En moyenne, un élève fait environ 4,33 trajets en bus par semaine.

Ex. 6 **(suite)**

(c) $N = 30$ est pair : la médiane est la moyenne des valeurs de rangs 15 et 16. D'après les effectifs cumulés, les rangs 11 à 19 valent 4.
Donc $Me = \frac{4+4}{2} = 4$. Au moins la moitié des élèves font au plus 4 trajets par semaine.

(d) $\frac{30}{4} = 7{,}5$: on prend le rang 8 (on arrondit à l'entier supérieur). Les rangs 1 à 10 valent 2, donc $Q_1 = 2$.
$\frac{3\times30}{4} = 22{,}5$: on prend le rang 23. Les rangs 20 à 26 valent 6, donc $Q_3 = 6$. Écart interquartile : $Q_3 - Q_1 = 6 - 2 = 4$ trajets.

(e) Maintenant $N = 31$ et la somme vaut $130 + 24 = 154$, donc $\overline{x} = \frac{154}{31} \approx 4{,}97$ trajets. $N = 31$ est impair : la médiane est la valeur de rang $\frac{31+1}{2} = 16$. Le rang 16 vaut encore 4, donc $Me = 4$. Une seule valeur extrême fait passer la moyenne d'environ 4,33 à environ 4,97 trajets, alors que la médiane ne bouge pas : la moyenne est sensible aux valeurs extrêmes, la médiane non.

Ex. 7 ★★★ **Quelle trottinette choisir ?**

(a) Série rangée : 28 ; 31 ; 34 ; 38 ; 41 ; 43 ; 45 ; 52. Somme : 312, donc $\overline{x} = \frac{312}{8} = 39$ km. $N = 8$ est pair : la médiane est la moyenne des valeurs de rangs 4 et 5, soit $Me = \frac{38+41}{2} = 39{,}5$ km. L'autonomie moyenne du modèle A est de 39 km et l'autonomie médiane de 39,5 km.

(b) $\frac{8}{4} = 2$: Q_1 est la valeur de rang 2, donc $Q_1 = 31$ km.
$\frac{3\times8}{4} = 6$: Q_3 est la valeur de rang 6, donc $Q_3 = 43$ km.
Écart interquartile : $43 - 31 = 12$ km. Étendue : $52 - 28 = 24$ km.

(c) $Q_3 = 42$ km signifie qu'au moins 75 % des trottinettes B testées ont une autonomie inférieure ou égale à 42 km, soit au moins $0{,}75 \times 8 = 6$ trottinettes sur 8.

(d) Modèle A : écart interquartile 12 km et étendue 24 km. Modèle B : écart interquartile $42 - 37 = 5$ km et étendue $45 - 35 = 10$ km. Les deux indicateurs sont plus petits pour B : les autonomies du modèle B sont beaucoup plus régulières que celles du modèle A.

(e) Modèle B : le minimum est de 35 km, donc toutes les trottinettes B testées couvrent le trajet. Modèle A : 28, 31 et 34 km sont inférieurs à 35 km, soit 3 trottinettes sur 8 (37,5 %) qui ne suffisent pas. Les moyennes sont proches (39 et 40 km), mais B est plus régulier. Je conseille à Léa le modèle B.

2. *Tableau croisé : fréquences marginales et conditionnelles*

Ex. 8 ★☆☆ **Fréquences lues dans un tableau croisé**

(a) $f(\text{natation}) = \frac{150}{250} = 0{,}6$.

(b) C'est la fréquence de la case : $\frac{66}{250} = 0{,}264 \approx 0{,}26$.

(c) Le dénominateur est le nombre de mineurs : $\frac{66}{120} = 0{,}55$.

(d) Le dénominateur est le nombre de grimpeurs : $\frac{66}{100} = 0{,}66$. Le numérateur est le même qu'à la question précédente, mais pas le dénominateur.

(e) $\frac{96}{130} \approx 0{,}74$.

Ex. 9 ★☆☆ **Tableau donné en pourcentages du total**

(a) On lit le total de la colonne Voiture : 46 %.

(b) On lit le total de la ligne Périphérie : 60 %.

(c) $\frac{12}{40} = 0{,}3$, soit 30 %.

(d) $\frac{12}{15} = 0{,}8$, soit 80 %.

(e) $\frac{36}{60} = 0{,}6$, soit 60 %.

(f) 21 % de 800 : $0{,}21 \times 800 = 168$ salariés.

Ex. 10 ★★☆ **Médiathèque : compléter puis calculer**

(a) Adultes : $480 - 200 = 280$. Pas de BD : $480 - 216 = 264$. Adultes BD : $216 - 140 = 76$. Enfants sans BD : $200 - 140 = 60$. Adultes sans BD : $280 - 76 = 204$. Vérification : $60 + 204 = 264$.

(b) $\frac{216}{480} = 0{,}45$: 45 % des inscrits empruntent des BD.

(c) Enfants : $\frac{140}{200} = 0{,}7$. Adultes : $\frac{76}{280} \approx 0{,}27$. Les enfants empruntent des BD beaucoup plus souvent que les adultes
(70 % contre environ 27 %).

(d) $\frac{140}{216} \approx 0{,}65$: environ 65 % des emprunteurs de BD sont des enfants. Il ne faut pas confondre ce résultat avec les 70 % d'enfants qui empruntent des BD.

Ex. 11 ★★☆ **Corriger un article de presse**

(a) $\frac{150}{600} = 0{,}25$ est la fréquence de la case « sportif et moins de 40 ans ». Parmi les moins de 40 ans, il faut diviser par 250 : $\frac{150}{250} = 0{,}6$.
Donc 60 % des moins de 40 ans font du sport.

(b) Non : les deux groupes n'ont pas la même taille, il faut comparer des fréquences. Pour les 40 ans et plus : $\frac{168}{350} = 0{,}48$, contre 0,6 pour les moins de 40 ans. Les moins de 40 ans sont donc plus souvent sportifs.

Ex. 11 (suite)

(c) Faux : il faut diviser par le nombre de sportifs, $150 + 168 = 318$. On obtient $\frac{150}{318} \approx 0{,}47$, soit moins de la moitié. Environ 47 % des sportifs ont moins de 40 ans. Il ne faut pas confondre ce résultat avec les 60 % de sportifs parmi les moins de 40 ans.

Ex. 12 ★★☆ Chauffage et étiquette énergie

(a) PAC : $\frac{64}{80} = 0{,}8$. Gaz : $\frac{45}{150} = 0{,}3$. Électrique : $\frac{21}{70} = 0{,}3$. 80 % des logements chauffés par PAC sont classés A à C, contre 30 % pour le gaz comme pour l'électrique.

(b) Faux : il y a plus de logements au gaz, donc comparer les effectifs n'a pas de sens. Les fréquences sont égales ($0{,}3$) : dans les deux cas, 30 % des logements sont classés A à C.

(c) $\frac{130}{300} \approx 0{,}43$: environ 43 % des logements sont classés A à C.

(d) Non : $\frac{0{,}8+0{,}3+0{,}3}{3} \approx 0{,}47$, ce qui est différent de 0,43, car les trois groupes n'ont pas la même taille. Moyenne pondérée par les effectifs : $\frac{80\times0{,}8+150\times0{,}3+70\times0{,}3}{300} = \frac{64+45+21}{300} = \frac{130}{300} \approx 0{,}43$. On retrouve la fréquence marginale : environ 43 % des logements sont classés A à C.

Ex. 13 ★★★ Internat : qui le choisit le plus souvent ?

(a) Parmi les 180 internes, 55 % sont des filles : $0{,}55 \times 180 = 99$ filles internes. Ces 99 filles représentent 36 % des filles, donc il y a $\frac{99}{0{,}36} = 275$ filles. Le lycée compte 99 filles internes et 275 filles.

(b) Garçons : $420 - 275 = 145$. Garçons internes : $180 - 99 = 81$. Filles externes : $275 - 99 = 176$. Garçons externes : $145 - 81 = 64$. Externes : $420 - 180 = 240$, et on vérifie que $176 + 64 = 240$. Tableau complété : Filles 99 | 176 | 275 ; Garçons 81 | 64 | 145 ; Total 180 | 240 | 420.

(c) La sous-population est celle des 145 garçons : $\frac{81}{145} \approx 0{,}56$. Environ 56 % des garçons sont internes.

(d) Non. Il compare des effectifs ($99 > 81$), alors que les deux groupes n'ont pas la même taille. Il faut comparer des fréquences conditionnelles : 36 % des filles sont internes, contre environ 56 % des garçons. Ce sont donc les garçons qui choisissent plus souvent l'internat. Les filles sont majoritaires à l'internat uniquement parce qu'il y a beaucoup plus de filles au lycée.

Ex. 13 (suite)

(e) Parmi les 240 externes : $\frac{176}{240} \approx 0{,}73$. Fréquence marginale des filles : $\frac{275}{420} \approx 0{,}65$. Comme $0{,}73 > 0{,}65$, les filles sont plus présentes chez les externes (environ 73 %) que dans l'ensemble du lycée (environ 65 %). Cela confirme qu'elles choisissent moins souvent l'internat.

Ex. 14 ★★★ Bilan de vue : enquête d'une mutuelle

(a) $0{,}30 \times 1\,200 = 360$ et $0{,}38 \times 1\,200 = 456$. 360 des 456 adhérents de 60 ans et plus ont fait un bilan.

(b) Les 10 % sont calculés sur l'effectif total : c'est la fréquence de la case « 18 à 39 ans et bilan fait », pas une fréquence parmi les 18 à 39 ans. Il y a $0{,}10 \times 1\,200 = 120$ bilans pour $0{,}32 \times 1\,200 = 384$ adhérents de cette tranche, donc $\frac{120}{384} \approx 0{,}31$. Phrase corrigée : « environ 31 % des 18 à 39 ans ont fait un bilan de vue. »

(c) 40 à 59 ans : $\frac{0{,}21\times1\,200}{0{,}30\times1\,200} = \frac{252}{360} = 0{,}70$. 60 ans et plus : $\frac{360}{456} \approx 0{,}79$. La fréquence passe d'environ 31 % à 70 %, puis à 79 % : ce sont les 60 ans et plus qui font le plus souvent un bilan de vue.

(d) Les bilans sont au nombre de $0{,}61 \times 1\,200 = 732$, donc $\frac{360}{732} \approx 0{,}49$: environ 49 % des adhérents ayant fait un bilan ont 60 ans ou plus. Le numérateur est toujours 360, mais le dénominateur change : 732 bilans ici, 456 adhérents de 60 ans et plus avant. « Parmi ceux qui ont fait un bilan, les seniors » n'est pas la même chose que « parmi les seniors, ceux qui ont fait un bilan ».

(e) Il faudrait $0{,}5 \times 384 = 192$ bilans, donc $192 - 120 = 72$ bilans en plus. Il y aurait alors $732 + 72 = 804$ bilans, et la fréquence marginale serait $\frac{804}{1\,200} = 0{,}67$. Il faut 72 bilans de plus chez les 18 à 39 ans, et 67 % des adhérents auraient alors fait un bilan.

3. Nuage de points et point moyen

Ex. 15 ★☆☆ Point moyen : questions flash

(a) $\overline{x} = \frac{1+3+5+7}{4} = 4$ et $\overline{y} = \frac{20+24+25+31}{4} = \frac{100}{4} = 25$, donc $G(4\,;\,25)$.

(b) $\overline{x} = \frac{2+5+8}{3} = 5$ et $\overline{y} = \frac{9+11+16}{3} = \frac{36}{3} = 12$, donc $G(5\,;\,12)$.

(c) On pondère par les effectifs : $\overline{x} = \frac{4\times3+8\times1}{4} = \frac{20}{4} = 5$ et $\overline{y} = \frac{5\times3+9\times1}{4} = \frac{24}{4} = 6$, donc $G(5\,;\,6)$.

(d) Rayan a fait la moyenne de toutes les valeurs mélangées. Il faut deux moyennes séparées : $\overline{x} = \frac{1+5}{2} = 3$ et $\overline{y} = \frac{9+13}{2} = 11$, donc $G(3\,;\,11)$.

(e) 2024 a pour rang $2024 - 2020 = 4$. Les rangs vont de 1 à 5, donc $\overline{x} = \frac{1+2+3+4+5}{5} = 3$.

Ex. 15 **(suite)**

(f) $\overline{y} = \frac{\text{somme des } y_i}{5} = 30$, donc la somme des y_i vaut $5 \times 30 = 150$.

Ex. 16 ★☆☆ **Lire un nuage de points**

(a) Le point d'abscisse 3 a pour ordonnée 50 : en semaine 3, Inès a parcouru 50 km.

(b) Le point d'ordonnée 60 a pour abscisse 4 : Inès a parcouru 60 km pendant la semaine 4.

(c) Maël a inversé les coordonnées. La semaine (variable x) se lit en abscisse et la distance (variable y) en ordonnée : le bon point est $(5\,;\,65)$.

(d) $70 - 40 = 30$: Inès a parcouru 30 km de plus en semaine 6 qu'en semaine 1.

(e) Les points d'ordonnée supérieure à 55 sont ceux des semaines 4, 5 et 6 (60, 65 et 70 km) : Inès a parcouru plus de 55 km pendant 3 semaines.

Ex. 17 ★★☆ **Révisions et notes : des points qui se répètent**

(a) Le nuage compte 4 points distincts, $(2\,;\,9)$, $(3\,;\,11)$, $(4\,;\,12)$ et $(5\,;\,14)$, et l'effectif total vaut $3 + 5 + 8 + 4 = 20$ élèves.

(b) $\overline{x} = \frac{2\times3+3\times5+4\times8+5\times4}{20} = \frac{73}{20} = 3{,}65$. En moyenne, ces élèves ont révisé 3,65 heures.

(c) $\overline{y} = \frac{9\times3+11\times5+12\times8+14\times4}{20} = \frac{234}{20} = 11{,}7$, donc $G(3{,}65\,;\,11{,}7)$. La note moyenne est de 11,7 sur 20.

(d) Noa a fait une moyenne simple des 4 notes. Or chaque note doit compter autant de fois qu'il y a d'élèves qui l'ont eue. La note 12, obtenue par 8 élèves, pèse plus lourd que les autres : la bonne moyenne est 11,7 sur 20.

(e) Non. Les abscisses des points du nuage sont des nombres entiers (2, 3, 4 ou 5), alors que celle de G vaut 3,65. Aucun élève n'a révisé 3,65 heures : G n'est pas un point du nuage.

Ex. 18 ★★☆ **Énergie solaire : rang de l'année et point moyen**

(a) $\overline{x} = \frac{1+2+3+4+5}{5} = 3$. L'axe des abscisses est gradué par le rang (de 1 à 5) et non par l'année : une abscisse de 2023 sortirait complètement du repère. Le rang 3 correspond bien à l'année 2023.

(b) $\overline{y} = \frac{120+132+141+155+167}{5} = \frac{715}{5} = 143$, donc $G(3\,;\,143)$. La production moyenne est de 143 MWh par an.

Ex. 18 **(suite)**

(c) Sur le nuage, le point de rang 3 est juste au-dessus de la graduation 140 : son ordonnée est 141 (valeur du tableau). Or $G(3\,;\,143)$ a pour ordonnée 143, et $143 \neq 141$. G n'est donc pas un point du nuage : il est un peu au-dessus du point de 2023. En général, le point moyen n'est pas un point du nuage.

(d) Inès obtient $\frac{1+2+3+4+5+120+132+141+155+167}{10} = \frac{730}{10} = 73$. Elle a tort : G a deux coordonnées, calculées séparément, $\overline{x} = 3$ et $\overline{y} = 143$. Son 73 mélange des rangs et des MWh et n'a pas de sens. Le point moyen est $G(3\,;\,143)$.

(e) 2026 a pour rang 6. $\overline{x} = \frac{1+2+3+4+5+6}{6} = 3{,}5$ et $\overline{y} = \frac{715+179}{6} = \frac{894}{6} = 149$, donc $G'(3{,}5\,;\,149)$: sur les six années, la production moyenne est de 149 MWh par an.

Ex. 19 ★★☆ **Transports : nuage allongé ou nuage courbé ?**

(a) Entre les mois 1 et 2 : $3 - 2 = 1$, soit 1 millier d'inscrits en plus. Entre les mois 5 et 6 : $28 - 16 = 12$, soit 12 milliers d'inscrits en plus.

(b) Les augmentations sont de plus en plus grandes : le nombre d'inscrits croît de plus en plus vite et le nuage A est courbé. Une droite résume mal cette tendance, donc un ajustement affine n'est pas pertinent.

(c) Le nuage B monte régulièrement (augmentations de 4, 3, 5, 3 et 4 centaines de passagers) : ses points sont à peu près alignés. Un ajustement affine est donc pertinent.

(d) Non. $\overline{y} = \frac{2+3+5+9+16+28}{6} = \frac{63}{6} = 10{,}5$, et non 15. Il faut faire la moyenne des six valeurs, pas seulement de la première et de la dernière : sur ces six mois, l'appli compte en moyenne 10,5 milliers d'inscrits.

Ex. 20 ★★★ **Chauffage et température : corriger puis ajuster**

(a) La moyenne de Jade est juste : $\frac{42+240}{12} = \frac{282}{12} = 23{,}5$. Mais elle mélange des températures et des consommations, alors qu'il faut deux moyennes séparées. $\overline{x} = \frac{6+2+10+4+12+8}{6} = \frac{42}{6} = 7$ et $\overline{y} = \frac{43+55+30+49+26+37}{6} = \frac{240}{6} = 40$. Le point moyen est donc $G(7\,;\,40)$: 7 °C et 40 kWh.

(b) Rangés : $(2\,;\,55)$, $(4\,;\,49)$, $(6\,;\,43)$, $(8\,;\,37)$, $(10\,;\,30)$, $(12\,;\,26)$. Les baisses sont de 6, 6, 6, 7 et 4 kWh : quand la température monte de 2 °C, la consommation baisse d'environ 6 kWh. Le nuage est allongé selon une droite qui descend.

Ex. 20 **(suite)**

(c) $-2{,}97 \times 7 + 60{,}8 = 40{,}01 \approx 40 = \overline{y}$: aux arrondis des coefficients près, $G(7\,;\,40)$ est bien sur la droite d'ajustement.

(d) $-2{,}97 \times 5 + 60{,}8 = 45{,}95 \approx 46$. Un jour à 5 °C, on peut estimer que la maison consomme environ 46 kWh. C'est une interpolation, car 5 °C est compris entre 2 °C et 12 °C, les températures extrêmes relevées.

(e) On résout $-2{,}97x + 60{,}8 = 35$: $x = \frac{35-60{,}8}{-2{,}97} = \frac{25{,}8}{2{,}97} \approx 8{,}7$. La consommation baisse quand la température monte : d'après cet ajustement, la maison consomme moins de 35 kWh dès qu'il fait plus de 8,7 °C environ.

Ex. 21 ★★★ **Club d'escalade : Mayer avec un effectif impair**

(a) La variable explicative est le temps : on la place en abscisse, et le nombre d'adhérents en ordonnée. En abscisse, on met le rang, pas l'année : 2025 a le rang 7. Donc $M_7(7\,;\,192)$, et non $(192\,;\,2025)$.

(b) Les hausses sont $140 - 132 = 8$, $154 - 140 = 14$, $161 - 154 = 7$, $171 - 161 = 10$, $184 - 171 = 13$ et $192 - 184 = 8$. Elles varient autour de 10 adhérents par an, sans accélérer ni ralentir. Le nuage est allongé selon une droite, donc un ajustement affine est pertinent.

(c) G_1 : $\frac{1+2+3}{3} = 2$ et $\frac{132+140+154}{3} = \frac{426}{3} = 142$, donc $G_1(2\,;\,142)$. G_2 : $\frac{4+5+6+7}{4} = 5{,}5$ et $\frac{161+171+184+192}{4} = \frac{708}{4} = 177$, donc $G_2(5{,}5\,;\,177)$. $a = \frac{177-142}{5{,}5-2} = \frac{35}{3{,}5} = 10$ et $b = 142 - 10 \times 2 = 122$. La droite de Mayer a pour équation $y = 10x + 122$.

(d) $\overline{x} = \frac{1+2+3+4+5+6+7}{7} = 4$ et $\overline{y} = \frac{1134}{7} = 162$, donc $G(4\,;\,162)$. $10 \times 4 + 122 = 162$: G est bien sur la droite. Adam a tort : le milieu de $[G_1G_2]$ est $\left(\frac{2+5{,}5}{2}\,;\,\frac{142+177}{2}\right) = (3{,}75\,;\,159{,}5)$, différent de G. Les deux groupes n'ont pas le même effectif (3 et 4).

(e) 2027 a le rang 9 : $10 \times 9 + 122 = 212$. On peut estimer que le club comptera environ 212 adhérents en 2027. C'est une extrapolation, car le rang 9 est en dehors des rangs observés (1 à 7). L'estimation est moins fiable : elle suppose que la hausse d'environ 10 adhérents par an continue.

4. *Droite de Mayer*

Ex. 22 ★☆☆ **Points moyens express**

(a) $\overline{x} = \frac{1+3+5+7}{4} = 4$ et $\overline{y} = \frac{12+15+19+22}{4} = \frac{68}{4} = 17$, donc $G(4\,;\,17)$.

(b) x : $\frac{1+2+6}{3} = 3$; y : $\frac{8+11+14}{3} = \frac{33}{3} = 11$, donc $G_1(3\,;\,11)$.

(c) Non. G_2 est un point moyen : x : $\frac{10+12+14+16}{4} = 13$; y : $\frac{40+46+43+51}{4} = \frac{180}{4} = 45$.
Donc $G_2(13\,;\,45)$. Tom a pris $\frac{40+51}{2} = 45{,}5$, qui n'est pas la moyenne des y.

(d) On range d'abord par x croissants : $(3\,;\,15)$, $(5\,;\,21)$ | $(7\,;\,26)$, $(9\,;\,23)$.
Donc $G_1\left(\frac{3+5}{2}\,;\,\frac{15+21}{2}\right) = G_1(4\,;\,18)$.

(e) $7 - 3 = 4$: le second groupe compte 4 points.

(f) Effectifs égaux : G est le milieu de $[G_1G_2]$, donc $G\left(\frac{3+9}{2}\,;\,\frac{12+20}{2}\right) = G(6\,;\,16)$.

Ex. 23 ★☆☆ **De G_1 et G_2 à l'équation**

(a) Non : il faut le même ordre en haut et en bas.
$a = \frac{17-5}{8-2} = \frac{12}{6} = 2$. Ali trouve $\frac{12}{-6} = -2$: sa droite descend, alors qu'elle doit monter.

(b) $a = \frac{28-40}{7-3} = \frac{-12}{4} = -3$.

(c) $b = 10 - 1{,}5 \times 4 = 4$.

(d) $a = \frac{47-32}{20-10} = 1{,}5$ et $b = 32 - 1{,}5 \times 10 = 17$: $y = 1{,}5x + 17$.

(e) $2 \times 5 + 3 = 13$: oui, la droite passe par G.

(f) $0{,}5 \times 8 + 6 = 10 \neq 11$: la droite ne passe pas par G, donc un calcul est faux.

(g) $a = \frac{9-4}{4-1} = \frac{5}{3}$ et $b = 4 - \frac{5}{3} \times 1 = \frac{7}{3}$: $y = \frac{5}{3}x + \frac{7}{3}$.

Ex. 24 ★★☆ **Chauffage et température**

(a) Série rangée : $(1\,;\,61)$, $(4\,;\,52)$, $(7\,;\,43)$ | $(8\,;\,38)$, $(10\,;\,33)$, $(12\,;\,29)$, $(14\,;\,24)$. Le premier groupe a 3 points, le second 4.

(b) $G_1\left(\frac{1+4+7}{3}\,;\,\frac{61+52+43}{3}\right) = G_1(4\,;\,52)$ et $G_2\left(\frac{8+10+12+14}{4}\,;\,\frac{38+33+29+24}{4}\right) = G_2(11\,;\,31)$.

(c) $a = \frac{31-52}{11-4} = \frac{-21}{7} = -3$ et $b = 52 - (-3) \times 4 = 64$. La droite de Mayer a pour équation $y = -3x + 64$.

(d) $\overline{x} = \frac{56}{7} = 8$ et $\overline{y} = \frac{280}{7} = 40$, donc $G(8\,;\,40)$. On a $-3 \times 8 + 64 = 40$: G est bien sur la droite, le contrôle est réussi.

Ex. 24 (suite)

(e) $-3 \times 9 + 64 = 37$. C'est une interpolation, car 9 est compris entre 1 et 14. On peut estimer la consommation à environ 37 kWh pour une journée à 9 °C.

Ex. 25 ★★☆ L'erreur de Lou

(a) Lou a pris les 3 premières lignes du tableau, dans l'ordre donné : $\frac{5+2+9}{3} = \frac{16}{3}$ et $\frac{50+58+42}{3} = 50$. Lou n'a pas rangé la série par x croissants avant de couper : ses groupes mélangent petites et grandes durées.

(b) Série rangée : $x = 2, 3, 4 \mid 5, 7, 9$.
$G_1\left(\frac{2+3+4}{3} ; \frac{58+55+52}{3}\right) = G_1(3 ; 55)$ et
$G_2\left(\frac{5+7+9}{3} ; \frac{50+46+42}{3}\right) = G_2(7 ; 46)$.

(c) $a = \frac{46-55}{7-3} = -2{,}25$ et $b = 55 + 2{,}25 \times 3 = 61{,}75$: la droite de Mayer est $y = -2{,}25x + 61{,}75$.

(d) $\overline{x} = \frac{30}{6} = 5$ et $\overline{y} = \frac{303}{6} = 50{,}5$, donc $G(5 ; 50{,}5)$. On a $-2{,}25 \times 5 + 61{,}75 = 50{,}5$: le contrôle est réussi.

(e) $-1{,}5 \times 5 + 58 = 50{,}5$: oui, la droite de Lou passe aussi par G. Toute droite (G_1G_2) passe par G, même avec des groupes mal formés. Le contrôle repère une erreur de calcul, mais il ne prouve pas que la série a été rangée.

Ex. 26 ★★☆ Mayer ou calculatrice ?

(a) $G_1\left(\frac{1+2+3+4}{4} ; \frac{12{,}4+13{,}1+14{,}5+15}{4}\right) = G_1(2{,}5 ; 13{,}75)$ et
$G_2\left(\frac{5+6+7+8}{4} ; \frac{16{,}2+16{,}8+18{,}1+18{,}7}{4}\right) = G_2(6{,}5 ; 17{,}45)$.

(b) $a = \frac{17{,}45-13{,}75}{6{,}5-2{,}5} = 0{,}925$ et $b = 13{,}75 - 0{,}925 \times 2{,}5 = 11{,}4375$: $y = 0{,}925x + 11{,}4375$. Contrôle : $G(4{,}5 ; 15{,}6)$ et $0{,}925 \times 4{,}5 + 11{,}4375 = 15{,}6$. La droite passe bien par G.

(c) Mayer : $0{,}925 \times 10 + 11{,}4375 = 20{,}6875$, soit environ 2 069 repas. Calculatrice : $0{,}919 \times 10 + 11{,}46 = 20{,}65$, soit environ 2 065 repas. C'est une extrapolation, car 10 est au-delà du mois 8.

(d) Les deux méthodes ne calculent pas la droite de la même façon : elles donnent des droites voisines, mais différentes. Les deux estimations ne diffèrent que d'environ 4 repas. Mais une extrapolation reste peu fiable si la tendance change après le mois 8.

Ex. 27 ★★★ Le dénivelé d'Inès

(a) Série rangée : x : 2 ; 3 ; 4 ; 6 ; 7 ; 9 ; 12 et y : 68 ; 71 ; 77 ; 84 ; 89 ; 96 ; 107. $G_1\left(\frac{2+3+4}{3} ; \frac{68+71+77}{3}\right) = G_1(3 ; 72)$ et $G_2\left(\frac{34}{4} ; \frac{84+89+96+107}{4}\right) = G_2(8{,}5 ; 94)$. $a = \frac{94-72}{8{,}5-3} = \frac{22}{5{,}5} = 4$ et $b = 72 - 4 \times 3 = 60$. La droite de Mayer est $y = 4x + 60$: chaque centaine de mètres de dénivelé ajoute environ 4 min.

Ex. 27 (suite)

(b) $\overline{x} = \frac{2+3+4+6+7+9+12}{7} = \frac{43}{7}$ et $\overline{y} = \frac{216+376}{7} = \frac{592}{7}$.
Contrôle : $4 \times \frac{43}{7} + 60 = \frac{172}{7} + \frac{420}{7} = \frac{592}{7}$, donc G est bien sur la droite. Avec les arrondis 6,14 et 84,57, on trouverait $4 \times 6{,}14 + 60 = 84{,}56$: le contrôle semblerait échouer. Il faut donc garder les valeurs exactes.

(c) Le milieu de $[G_1G_2]$ a pour abscisse $\frac{3+8{,}5}{2} = 5{,}75$, alors que $\overline{x} = \frac{43}{7} \approx 6{,}14$. Léa a donc tort. Les deux groupes n'ont pas le même effectif (3 et 4) : $\overline{x} = \frac{3\times3+4\times8{,}5}{7} = \frac{43}{7}$. G est donc plus proche de G_2, mais il reste sur la droite (G_1G_2).

(d) 1 500 m correspond à $x = 15$, et $4 \times 15 + 60 = 120$. On estime la durée à 120 min, soit 2 h. C'est une extrapolation, car $15 > 12$ (le plus grand dénivelé relevé). L'estimation est peu fiable : la fatigue peut ralentir Inès plus que prévu sur un si grand dénivelé.

(e) 1 h 40 min = 100 min. On résout $4x + 60 = 100$, d'où $x = \frac{40}{4} = 10$, et $4x + 60 \leqslant 100$ équivaut à $x \leqslant 10$. Inès peut choisir au plus 1 000 m de dénivelé environ. C'est une interpolation ($2 \leqslant 10 \leqslant 12$), donc l'estimation est assez fiable.

Ex. 28 ★★★ Vélos en libre-service

(a) Les points sont presque alignés et montent régulièrement, d'environ 0,6 à 0,9 millier par an. Le nuage n'est pas courbé, donc un ajustement affine est pertinent.

(b) Les rangs sont déjà rangés par ordre croissant. G_1 : $\frac{1+2+3+4}{4} = 2{,}5$ et $\frac{4{,}2+4{,}9+5{,}5+6{,}4}{4} = \frac{21}{4} = 5{,}25$, donc $G_1(2{,}5 ; 5{,}25)$. G_2 : $\frac{5+6+7+8}{4} = 6{,}5$ et $\frac{7+7{,}6+8{,}5+9{,}1}{4} = \frac{32{,}2}{4} = 8{,}05$, donc $G_2(6{,}5 ; 8{,}05)$. $a = \frac{8{,}05-5{,}25}{6{,}5-2{,}5} = \frac{2{,}8}{4} = 0{,}7$ et $b = 5{,}25 - 0{,}7 \times 2{,}5 = 3{,}5$. La droite de Mayer a pour équation $y = 0{,}7x + 3{,}5$.

(c) $\overline{x} = \frac{36}{8} = 4{,}5$ et $\overline{y} = \frac{21+32{,}2}{8} = \frac{53{,}2}{8} = 6{,}65$, donc $G(4{,}5 ; 6{,}65)$. Premier contrôle : $0{,}7 \times 4{,}5 + 3{,}5 = 6{,}65$, donc G est sur la droite. Second contrôle : les groupes ont le même effectif, et $\frac{2{,}5+6{,}5}{2} = 4{,}5$, $\frac{5{,}25+8{,}05}{2} = 6{,}65$, donc G est le milieu de $[G_1G_2]$.

(d) Avec Mayer, $0{,}7x + 3{,}5 > 12$ donne $x > \frac{8{,}5}{0{,}7} \approx 12{,}14$. Avec la calculatrice, $0{,}705x + 3{,}479 > 12$ donne $x > \frac{8{,}521}{0{,}705} \approx 12{,}09$. Dans les deux cas, le premier rang entier est 13, donc l'année 2030. Les droites sont voisines mais différentes, et elles donnent la même conclusion. Mais c'est une extrapolation loin des données ($13 > 8$) : la tendance peut changer d'ici 2030, donc la prévision reste fragile.

Ex. 28 **(suite)**

(e) Il a tort. Le nuage relie seulement le rang de l'année au nombre d'abonnés : les pistes cyclables n'y sont même pas mesurées. Même un lien statistique ne prouverait pas une causalité : d'autres facteurs varient avec le temps (prix du carburant, habitudes, nombre d'habitants, nombre de stations) et peuvent expliquer la hausse des abonnements.

5. *Droite d'ajustement : estimer, interpoler, extrapoler*

Ex. 29 ★☆☆ **Automatismes : remplacer ou résoudre**

(a) $y = 1{,}8 \times 3 + 12{,}5 = 5{,}4 + 12{,}5 = 17{,}9$, soit environ 1 790 visiteurs au mois 3.

(b) 3 est compris entre 1 et 6, les mois observés : c'est une interpolation.

(c) $y = 1{,}8 \times 9 + 12{,}5 = 28{,}7$ centaines, soit environ 2 870 visiteurs au mois 9 (extrapolation, car $9 > 6$).

(d) 2 150 visiteurs correspondent à $y = 21{,}5$. On résout $1{,}8x + 12{,}5 = 21{,}5$, donc $1{,}8x = 9$ et $x = \frac{9}{1{,}8} = 5$. La droite prévoit 2 150 visiteurs au mois 5.

(e) Chaque mois, la fréquentation augmente en moyenne de 1,8 centaine, soit 180 visiteurs.

Ex. 30 ★☆☆ **Automatismes : le rang, pas l'année**

(a) $2024 - 2018 = 6$: l'année 2024 a le rang 6.

(b) $2030 - 2018 = 12$: l'année 2030 a le rang 12.

(c) $2018 + 10 = 2028$: le rang 10 correspond à l'année 2028.

(d) $y = 0{,}35 \times 12 + 4{,}2 = 8{,}4$ centaines, soit environ 840 vélos en 2030 (extrapolation, car $12 > 7$).

(e) Sam trouve 714,7 centaines, soit 71 470 vélos : c'est absurde pour un magasin qui en vend quelques centaines par an. Il a remplacé x par l'année au lieu du rang 12.

Ex. 31 ★★☆ **Consommation d'eau : quand passer sous un seuil ?**

(a) $\overline{x} = \frac{1+2+\cdots+7}{7} = 4$ et $\overline{y} = \frac{151+149+147+143+140+137+134}{7} = \frac{1\,001}{7} = 143$, donc $G(4\,;\,143)$. On calcule $-2{,}93 \times 4 + 154{,}71 = 142{,}99 \approx 143$: la droite passe presque par G (l'écart vient de l'arrondi des coefficients).

Ex. 31 **(suite)**

(b) 2027 a le rang 10 : $y = -2{,}93 \times 10 + 154{,}71 = 125{,}41$. La droite prévoit une consommation d'environ 125,4 L par jour et par habitant en 2027. C'est une extrapolation, car $10 > 7$.

(c) On résout $-2{,}93x + 154{,}71 = 120$: $-2{,}93x = -34{,}71$, donc $x = \frac{34{,}71}{2{,}93} \approx 11{,}85$. La consommation baisse, elle passe donc sous 120 L après ce rang, c'est-à-dire au rang entier supérieur 12, soit l'année $2017 + 12 = 2029$. La commune atteindrait son objectif en 2029 (par extrapolation).

(d) 2070 a le rang 53 : $y = -2{,}93 \times 53 + 154{,}71 = -0{,}58$. Une consommation négative est impossible : cette extrapolation est trop lointaine (46 ans après les données), la baisse ne peut pas rester affine. Au mieux, une telle estimation donne un ordre de grandeur, ici elle est fausse.

Ex. 32 ★★☆ **Course à pied : entraînement et temps sur 10 km**

(a) $\overline{x} = \frac{10+15+20+25+30+50}{6} = \frac{150}{6} = 25$ et $\overline{y} = \frac{62+58+55+51+49+40}{6} = \frac{315}{6} = 52{,}5$, donc $G(25\,;\,52{,}5)$. On calcule $-0{,}54 \times 25 + 66 = 52{,}5$: la droite passe bien par G.

(b) $y = -0{,}54 \times 35 + 66 = 47{,}1$ min, soit 47 min 6 s, car 0,1 min = 6 s. C'est une interpolation, car 35 est compris entre 10 et 50. On estime que Maëlle court le 10 km en environ 47 min 6 s.

(c) Estimation : $-0{,}54 \times 30 + 66 = 49{,}8$ min. Écart : $49{,}8 - 49 = 0{,}8$ min, soit 48 s. Ce coureur est 48 s plus rapide que ce que prévoit la droite.

(d) Ici, 45 est un temps, donc une valeur de y : il faut résoudre l'équation, pas remplacer x. $-0{,}54x + 66 = 45$ donne $0{,}54x = 21$, donc $x = \frac{21}{0{,}54} \approx 38{,}9$. Tom devrait courir environ 38,9 km par semaine (interpolation).

Ex. 33 ★★☆ **Bornes de recharge : corrige la copie de Kenza**

(a) Comme $x = 1$ pour 2019, l'année 2023 a le rang $2023 - 2018 = 5$, et non 4. On calcule $y = 6{,}63 \times 5 + 6{,}8 = 39{,}95$, soit environ 40 bornes. Le nuage montre 40 bornes en 2023 : l'écart n'est que de 0,05 borne. La droite estime donc très bien l'année 2023.

(b) 2028 a le rang $2028 - 2018 = 10$. On calcule $y = 6{,}63 \times 10 + 6{,}8 = 73{,}1$. La ville aurait environ 73 bornes en 2028. C'est une extrapolation, car $10 > 6$.

Ex. 33 **(suite)**

(c) b est l'estimation pour $x = 0$, c'est-à-dire l'année 2018, et non 2019. Cette valeur est déjà une extrapolation. En 2019, le nuage montre environ 13 bornes. Il faut écrire : « la droite estime à environ 7 le nombre de bornes en 2018 ».

(d) x est un rang d'année : $a = 6{,}63$ signifie que la ville gagne en moyenne environ 6,63 bornes par an, et non par mois
(environ 7 bornes de plus chaque année).

Ex. 34 ★★★ **Vélos en libre-service : de la droite à la décision**

(a) 2021 correspond au rang $x = 4$. On calcule $y = 1{,}675 \times 4 + 10{,}7 = 17{,}4$, soit 1 740 abonnés. La valeur observée est 17,1 centaines, soit 1 710 abonnés. L'écart vaut $17{,}4 - 17{,}1 = 0{,}3$ centaine. En 2021, la droite surestime donc le nombre d'abonnés de 30 abonnés.

(b) 2026 correspond au rang $x = 2026 - 2017 = 9$, et non à $x = 2026$. On calcule $y = 1{,}675 \times 9 + 10{,}7 = 25{,}775$. La droite prévoit environ 2 578 abonnés en 2026. Comme 9 est en dehors de la plage observée $[1\,;\,7]$, c'est une extrapolation.

(c) On veut $y \geqslant 33$. On résout $1{,}675x + 10{,}7 \geqslant 33$, soit $1{,}675x \geqslant 22{,}3$ et $x \geqslant \frac{22{,}3}{1{,}675} \approx 13{,}31$. Le rang est entier, donc on prend l'entier supérieur : $x = 14$. Vérification : pour $x = 13$, $y = 32{,}475 < 33$; pour $x = 14$, $y = 34{,}15 \geqslant 33$. Le rang 14 correspond à $2017 + 14 = 2031$. Par extrapolation, l'objectif de 3 300 abonnés serait atteint en 2031.

(d) 2050 correspond au rang $x = 2050 - 2017 = 33$. On calcule $y = 1{,}675 \times 33 + 10{,}7 = 65{,}975$, soit environ 6 600 abonnés : le calcul va bien dans le sens de l'élu. Mais c'est une extrapolation très lointaine, à 26 ans de la dernière donnée. Elle suppose que la hausse reste la même jusqu'en 2050, alors qu'une saturation est probable (taille de la population, nombre de vélos disponibles). Ce résultat n'est qu'un ordre de grandeur peu fiable : l'affirmation de l'élu n'est pas justifiée.

Ex. 35 ★★★ **Chauffage et température : jusqu'où croire la droite ?**

(a) $a = -3{,}73$: quand la température moyenne augmente de 1 °C, la consommation baisse en moyenne de 3,73 kWh par jour. Pour $x = 0$, on calcule $y = -3{,}73 \times 0 + 87{,}91 \approx 87{,}9$ kWh par jour. La valeur observée est 88 kWh par jour, soit un écart d'environ 0,1 kWh. Ici, 0 est dans la plage observée $[-2; 10]$: l'ordonnée à l'origine a un sens concret et la droite estime très bien la consommation à 0 °C.

(b) On calcule $y = -3{,}73 \times 6{,}5 + 87{,}91 = 63{,}665 \approx 63{,}7$. Pour une semaine à 6,5 °C, la consommation estimée est d'environ 63,7 kWh par jour. Comme 6,5 est compris entre -2 et 10, c'est une interpolation.

(c) On cherche x : on résout l'inéquation, on ne remplace pas. $-3{,}73x + 87{,}91 \leqslant 60$, soit $-3{,}73x \leqslant -27{,}91$. On divise par un nombre négatif, donc le sens de l'inégalité change : $x \geqslant \frac{27{,}91}{3{,}73} \approx 7{,}5$. Lina consomme au plus 60 kWh par jour quand la température moyenne est d'au moins 7,5 °C environ. Cette valeur est comprise entre -2 et 10 : c'est une interpolation.

(d) On calcule $y = -3{,}73 \times (-10) + 87{,}91 = 37{,}3 + 87{,}91 = 125{,}21 \approx 125{,}2$ kWh par jour : son calcul est juste. Mais $-10 < -2$: c'est une extrapolation, en dehors de la plage observée $[-2\,;\,10]$. Rien ne garantit que la consommation reste affine par grand froid (puissance maximale du chauffage, pertes plus fortes). 125 kWh par jour n'est qu'un ordre de grandeur : Lina a tort d'affirmer que c'est certain.

6. *Choisir un ajustement, corrélation et causalité*

Ex. 36 ★☆☆ **Questions flash : indicateurs et point moyen**

(a) Série rangée : 9 ; 11 ; 12 ; 15 ; 18. Il y a 5 valeurs, la médiane est la 3e : 12.

(b) $\overline{x} = \frac{2+4+6+8}{4} = 5$ et $\overline{y} = \frac{30+26+21+15}{4} = \frac{92}{4} = 23$, donc $G(5\,;\,23)$.

(c) $-1{,}5 \times 6 + 30 = -9 + 30 = 21$: les coordonnées de G vérifient l'équation, donc la droite passe par G.

(d) $n = 8$: Q_1 est la 2e valeur, $Q_1 = 7$; Q_3 est la 6e valeur, $Q_3 = 13$. L'écart interquartile vaut $13 - 7 = 6$.

(e) $\frac{10 \times 12 + 15 \times 8}{25} = \frac{240}{25} = 9{,}6$: la moyenne du groupe est 9,6.

Ex. 36 **(suite)**

(f) 4,5 est compris entre 1 et 8 : c'est une interpolation. $10 > 8$ est hors de la plage observée : c'est une extrapolation.

Ex. 37 ★☆☆ **Vélo et arrêts maladie : tableau croisé**

(a) Autre transport et au moins un arrêt : $140 - 84 = 56$. Total sans arrêt : $48 + 84 = 132$.

(b) $\frac{48}{200} = 0{,}24$.

(c) $\frac{60}{200} = 0{,}3$.

(d) $\frac{12}{60} = 0{,}2$.

(e) $\frac{56}{140} = 0{,}4$: 40 % des salariés qui utilisent un autre transport ont eu au moins un arrêt, contre 20 % des cyclistes, soit deux fois plus.

(f) $\frac{48}{132} \approx 0{,}36$.

(g) Faux. Le tableau montre un lien (0,2 contre 0,4), pas une cause. D'autres explications sont possibles. Un salarié souvent malade prend moins le vélo : la cause irait alors dans l'autre sens. Les salariés plus âgés viennent moins à vélo et sont plus souvent arrêtés : l'âge est une variable cachée.

Ex. 38 ★★☆ **Glaces et coups de soleil à Royan**

(a) $\overline{x} = \frac{4+6+9+12+14+15}{6} = \frac{60}{6} = 10$ et $\overline{y} = \frac{3+5+8+10+13+15}{6} = \frac{54}{6} = 9$, donc $G(10\,;\,9)$.

(b) La calculatrice donne $a \approx 1{,}03$ et $b \approx -1{,}31$, soit $y = 1{,}03x - 1{,}31$. Contrôle : $1{,}03 \times 10 - 1{,}31 = 8{,}99 \approx 9$, donc la droite passe par G, aux arrondis près.

(c) 1 100 glaces donnent $x = 11$: $y = 1{,}03 \times 11 - 1{,}31 = 10{,}02$, soit environ 10 consultations dans la semaine. Comme $4 \leqslant 11 \leqslant 15$, c'est une interpolation.

(d) Le titre confond lien et cause. La température augmente de 21 °C à 33 °C en même temps que les deux séries : la chaleur et le soleil font vendre des glaces et donnent des coups de soleil. C'est une variable cachée. Le nuage ne prouve donc pas que les glaces provoquent les coups de soleil.

Ex. 39 ★★☆ **Une citerne pendant la sécheresse**

(a) Le point $(4\,;\,20)$: les autres volumes baissent régulièrement d'environ 4 hL par semaine, et 20 hL est très en dessous de cette tendance. Ce point est aberrant.

Ex. 39 **(suite)**

(b) Avec 20 : $y = -4{,}26x + 50{,}73$. Avec 37 : $y = -3{,}77x + 51{,}87$. Une seule valeur aberrante change nettement la pente : elle déplace la droite. Après correction, le volume baisse en moyenne de 3,77 hL par semaine.

(c) $y = -3{,}77 \times 8 + 51{,}87 = 21{,}71$, soit environ 21,7 hL en semaine 8. C'est une extrapolation, car $8 > 6$.

(d) Avec la droite non corrigée : $y = -4{,}26 \times 8 + 50{,}73 = 16{,}65$, soit environ 16,7 hL. L'erreur de lecture aurait fait sous-estimer le volume de $21{,}71 - 16{,}65 = 5{,}06$ hL, soit environ 5 hL : une seule valeur aberrante fausse nettement l'estimation.

(e) $-3{,}77x + 51{,}87 = 15$ donne $x = \frac{51{,}87-15}{3{,}77} = \frac{36{,}87}{3{,}77} \approx 9{,}78$. Contrôle : 17,94 hL en semaine 9 et 14,17 hL en semaine 10. D'après la droite, Inès devra acheter de l'eau à partir de la semaine 10. C'est une extrapolation : une pluie ou une autre consommation changerait cette date.

Ex. 40 ★★☆ **Bornes de recharge : deux droites pour un nuage courbé**

(a) Les hausses annuelles grandissent $(2\,;\,4\,;\,6\,;\,9\,;\,13)$: le nuage est courbé vers le haut. Un ajustement affine n'est donc pas vraiment pertinent sur les six années.

(b) Écarts : $10 - 5{,}86 = 4{,}14$; $12 - 12{,}52 = -0{,}52$; $16 - 19{,}18 = -3{,}18$; $22 - 25{,}84 = -3{,}84$; $31 - 32{,}5 = -1{,}5$; $44 - 39{,}16 = 4{,}84$. Les signes sont positifs aux extrémités et négatifs au milieu : la droite ne suit pas la courbure du nuage.

(c) D_1 : $4{,}14 + 0{,}52 + 3{,}18 + 3{,}84 + 1{,}5 + 4{,}84 = 18{,}02$. D_2 donne -21 ; -8 ; 5 ; 18 ; 31 ; 44, d'où les écarts $31 + 20 + 11 + 4 + 0 + 0 = 66$. Sur l'ensemble des points, D_1 est donc la plus proche.

(d) D_1 : $6{,}66 \times 7 - 0{,}8 = 45{,}82$, soit environ 46 bornes. D_2 : $13 \times 7 - 34 = 57$ bornes. D_1 prévoit seulement 2 bornes de plus qu'en année 6 alors que la hausse s'accélère. D_2, qui suit les dernières années, paraît plus vraisemblable, mais cela reste une extrapolation.

Ex. 41 ★★★ **Neige au col : quelle droite, et jusqu'à quand ?**

(a) $\overline{x} = \frac{1+2+3+4+5}{5} = 3$ et $\overline{y} = \frac{92+80+71+58+49}{5} = \frac{350}{5} = 70$, donc $G(3\,;\,70)$. Pour D_1 : $-11 \times 3 + 105 = 72 \neq 70$, donc D_1 ne passe pas par G. Pour D_2 : $-10{,}8 \times 3 + 102{,}4 = 70$, donc D_2 passe par G.

Ex. 41 **(suite)**

(b) D_1 donne 94 ; 83 ; 72 ; 61 ; 50. Les écarts valent -2 ; -3 ; -1 ; -3 ; -1, et la somme des carrés vaut $4+9+1+9+1=24$. D_2 donne 91,6 ; 80,8 ; 70 ; 59,2 ; 48,4. Les écarts valent 0,4 ; $-0{,}8$; 1 ; $-1{,}2$; 0,6, et la somme des carrés vaut $0{,}16+0{,}64+1+1{,}44+0{,}36=3{,}6$. Comme $3{,}6<24$, D_2 est la plus proche des points : Maëlle retient D_2.

(c) $y=-10{,}8\times 7+102{,}4=26{,}8$. En semaine 7, il resterait environ 26,8 cm de neige. Comme $7>5$, en dehors de la plage observée, c'est une extrapolation.

(d) $y=-10{,}8\times 12+102{,}4=-27{,}2$. Une hauteur de $-27{,}2$ cm est impossible. Une droite décroissante finit toujours par donner des valeurs négatives : le modèle ne vaut que sur une plage limitée, proche des semaines observées.

(e) Avec D_2 : $-10{,}8x+102{,}4=0$, donc $10{,}8x=102{,}4$ et $x=\frac{102{,}4}{10{,}8}\approx 9{,}48$. D'après la droite, la neige disparaîtrait vers la semaine 9 ou 10. C'est une extrapolation assez lointaine : la fonte peut accélérer, avec un soleil plus chaud, ou ralentir, avec une nouvelle chute de neige. Ce n'est qu'un ordre de grandeur.

Ex. 42 ★★★ **« La musique fait monter les notes » : vraiment ?**

(a) Joue d'un instrument, moins de 12 : $80-48=32$. N'en joue pas, au moins 12 : $116-48=68$. Total des non-musiciens : $68+102=170$. Total des moins de 12 : $250-116=134$. On contrôle : $32+102=134$ et $80+170=250$.

(b) Parmi les 170 élèves qui ne jouent d'aucun instrument, 68 ont au moins 12 : $\frac{68}{170}=0{,}4$. 40 % de ces élèves ont au moins 12 de moyenne.

(c) Parmi les 116 élèves ayant au moins 12, 48 jouent d'un instrument : $\frac{48}{116}\approx 0{,}41$. Environ 41 % des bons élèves sont musiciens, et non 60 %. Le 60 % vient de $\frac{48}{80}=0{,}6$: c'est la part des bons élèves parmi les musiciens, contre 40 % chez les autres. Le journal a confondu deux fréquences conditionnelles.

(d) Non. Le tableau montre un lien statistique entre les deux caractères, pas une causalité. Une variable cachée peut expliquer le lien : une famille qui peut payer des cours de musique peut aussi aider aux devoirs. Un élève organisé peut aussi trouver le temps de faire les deux. Pour établir une cause, il faudrait une expérience contrôlée.

Ex. 42 **(suite)**

(e) Par exemple : « Dans ce lycée, 60 % des musiciens ont au moins 12 de moyenne, contre 40 % des autres élèves ». Ce titre décrit le lien observé sans affirmer que la musique en est la cause.