Statistiques à deux variables — Première générale

20 exercices corrigés · mathlvl.fr

Énoncés et corrigés détaillés

Présentation

Deux variables mesurées sur les mêmes individus donnent un nuage de points, et la question est toujours la même : ce nuage est-il assez allongé pour qu'une droite le résume, et que vaut alors la prévision qu'on en tire ? Ces 20 exercices installent le point moyen, l'ajustement affine — par les points extrêmes, par la méthode de Mayer, puis par les moindres carrés à la calculatrice —, les résidus et le coefficient de corrélation. La seconde moitié est consacrée à ce qui fait la différence en devoir : savoir linéariser par changement de variable, repérer un point aberrant, et surtout critiquer une prévision. Un coefficient de corrélation proche de 1 ne rend pas un modèle vrai, et une extrapolation lointaine peut annoncer 137 % de foyers équipés.

Version en ligne, avec les corrections dépliables :
mathlvl.fr/exercices-corriges/premiere-generale/statistiques-deux-variables

Exercice 1 – Point moyen d'un nuage

On a relevé, pour cinq magasins, le nombre x de vendeurs et le chiffre d'affaires y en milliers d'euros :

x : 1 ; 2 ; 3 ; 4 ; 5 y : 3 ; 5 ; 4 ; 8 ; 10

Calcule les coordonnées du point moyen G de ce nuage.

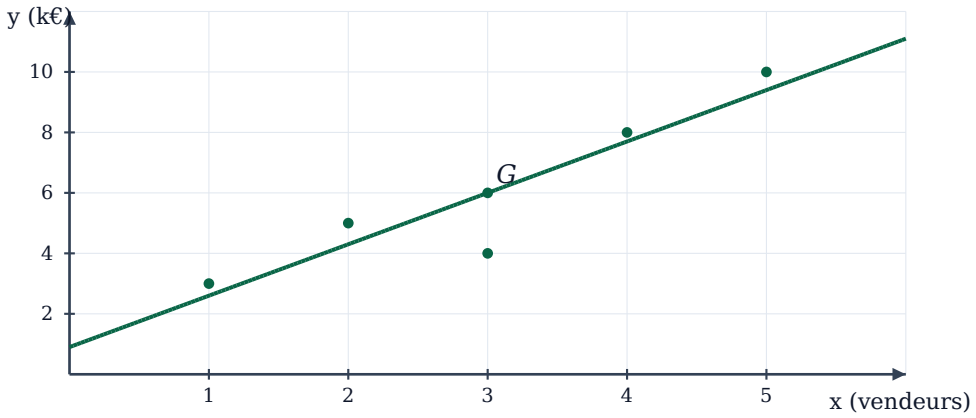
Exercice 2 – Décrire un nuage de points

Le nuage de la série précédente est représenté ci-dessous, avec sa droite d'ajustement.

1. La corrélation entre x et y est-elle positive ou négative ? Justifie.

2. Le nuage est-il bien aligné ? Quel point s'écarte le plus de la droite ?

3. Le point moyen $G(3\,;6)$ appartient-il à la droite ?



mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Exercice 3 – Moyennes d'une seconde série

Une entreprise relève sa dépense de publicité x (en milliers d'euros) et ses ventes y (en milliers d'articles) :

x : 0 ; 2 ; 4 ; 6 ; 8 y : 12 ; 15 ; 19 ; 22 ; 27

Calcule $\overline{x}$, $\overline{y}$ et les coordonnées du point moyen.

Exercice 4 – Utiliser une droite d'ajustement

Pour la série précédente, la calculatrice donne la droite d'ajustement $y = 1{,}85x + 11{,}6$.

1. Vérifie que cette droite passe par le point moyen $G(4\,;19)$.

2. Estime les ventes pour une dépense de 10 milliers d'euros.

3. Interprète le coefficient 1,85.

Exercice 5 – Interpolation ou extrapolation

Toujours avec la droite $y = 1{,}85x + 11{,}6$, obtenue à partir de dépenses comprises entre 0 et 8 milliers d'euros, on effectue trois estimations : pour $x = 5$, pour $x = 12$ et pour $x = 40$.

1. Calcule les trois valeurs estimées.

2. Pour chacune, dis s'il s'agit d'une interpolation ou d'une extrapolation.

3. Laquelle est la moins fiable, et pourquoi ?

Exercice 6 – Ajustement par les points extrêmes

On reprend la série des magasins : x : 1 ; 2 ; 3 ; 4 ; 5 et y : 3 ; 5 ; 4 ; 8 ; 10.

1. Détermine l'équation de la droite passant par le premier et le dernier point du nuage.

2. Cette droite passe-t-elle par le point moyen $G(3\,;6)$?

3. Que penser de cette méthode ?

Exercice 7 – Droite des moindres carrés

Pour la même série (x : 1 ; 2 ; 3 ; 4 ; 5 et y : 3 ; 5 ; 4 ; 8 ; 10), la calculatrice affiche l'ajustement affine $y = 1{,}7x + 0{,}9$.

1. Vérifie que cette droite passe par $G(3\,;6)$.

2. Compare-la à la droite des points extrêmes $y = 1{,}75x + 1{,}25$ pour $x = 2$ et $x = 5$.

3. Estime le chiffre d'affaires d'un magasin de 7 vendeurs.

Exercice 8 – Calculer des résidus

Avec l'ajustement $y = 1{,}7x + 0{,}9$, calcule pour chacun des cinq magasins la valeur estimée, puis le résidu (valeur réelle moins valeur estimée). Que vaut la somme des résidus ?

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Exercice 9 – Interpréter un coefficient de corrélation

Quatre études fournissent les coefficients de corrélation linéaire suivants. Pour chacun, indique le sens et la force de la liaison, et dis si un ajustement affine est pertinent.

1. $r = 0{,}996$ (dépense publicitaire et ventes)

2. $r = 0{,}92$ (nombre de vendeurs et chiffre d'affaires)

3. $r = -0{,}85$ (prix d'un produit et quantité vendue)

4. $r = 0{,}15$ (pointure et note en mathématiques)

Exercice 10 – Changer l'origine des années

Le chiffre d'affaires d'une société, en milliers d'euros, est relevé chaque année :

année : 2019 ; 2020 ; 2021 ; 2022 ; 2023 ; 2024 y : 242 ; 256 ; 279 ; 298 ; 313 ; 337

On pose x = année $-$ 2019.

1. Donne les valeurs de x et explique l'intérêt de ce changement.

2. Calcule $\overline{x}$ et $\overline{y}$.

3. La calculatrice donne $y = 19x + 240$. Vérifie qu'elle passe par le point moyen.

Exercice 11 – Méthode de Mayer

On considère la série :

x : 1 ; 2 ; 3 ; 4 ; 5 ; 6 y : 3 ; 5 ; 7 ; 10 ; 11 ; 12

Détermine la droite d'ajustement par la méthode de Mayer : on partage le nuage en deux sous-nuages de même effectif, on calcule le point moyen de chacun, et la droite passe par ces deux points.

Exercice 12 – Comparer deux ajustements

Pour la série précédente, la calculatrice donne la droite des moindres carrés $y \approx 1{,}886x + 1{,}4$, alors que la méthode de Mayer donnait $y = 2x + 1$.

1. Compare les estimations des deux modèles pour $x = 3$ et $x = 6$.

2. Compare-les pour $x = 10$.

3. Que conclure ?

Exercice 13 – Prévision et seuil

On garde le modèle de Mayer $y = 2x + 1$, où x est le rang du mois et y le nombre de commandes (en centaines).

1. Estime le nombre de commandes au rang 8.

2. À partir de quel rang le modèle prévoit-il plus de 25 centaines de commandes ?

3. À quelle date cela correspond-il si le rang 1 désigne janvier 2026 ?

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Exercice 14 – Linéariser par changement de variable

On relève la série suivante, dont le nuage est nettement courbé :

x : 1 ; 2 ; 3 ; 4 y : 2 ; 1 ; 0,667 ; 0,5

On pose $z = \dfrac{1}{y}$.

1. Calcule les valeurs de z et montre qu'elles sont alignées avec x.

2. Détermine la relation entre z et x, puis entre y et x.

3. Estime y pour $x = 10$.

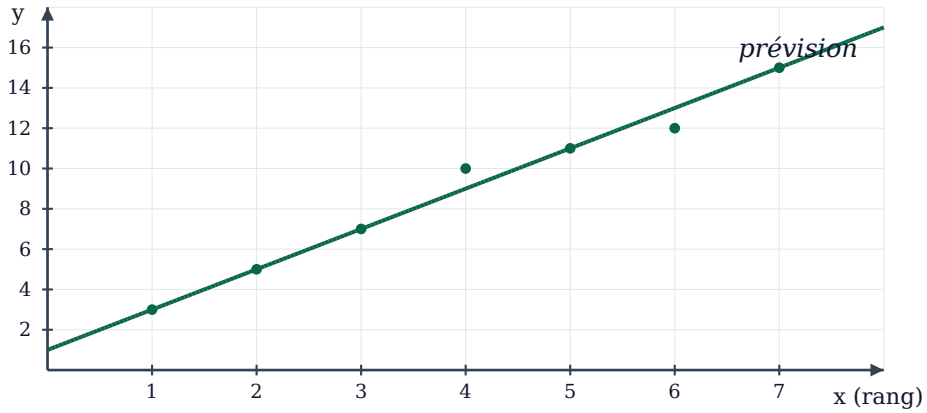
Exercice 15 – Lecture graphique d'un ajustement

Le nuage de la série (x : 1 à 6 ; y : 3 ; 5 ; 7 ; 10 ; 11 ; 12) est tracé ci-dessous avec la droite de Mayer.

1. Lis graphiquement l'ordonnée à l'origine et le coefficient directeur de la droite.

2. Estime graphiquement y pour $x = 7$, puis vérifie par le calcul.

3. Quels points du nuage sont au-dessus de la droite ?



Exercice 16 – Effet d'un point aberrant

Une série de six relevés donne : x : 1 ; 2 ; 3 ; 4 ; 5 ; 6 et y : 5 ; 8 ; 11 ; 14 ; 17 ; 20. La calculatrice donne l'ajustement $y = 3x + 2$ avec $r = 1$.

À la saisie, la dernière valeur est tapée par erreur : 2 au lieu de 20. La calculatrice donne alors $y \approx 0{,}43x + 8$ avec $r \approx 0{,}14$.

1. Vérifie que la série initiale est parfaitement alignée.

2. Compare les deux ajustements et les deux coefficients de corrélation.

3. Que faire face à un point qui s'écarte fortement des autres ?

Exercice 17 – Corrélation forte, modèle inadapté

Le nombre d'abonnés d'un service, en milliers, évolue ainsi :

x (années) : 0 ; 1 ; 2 ; 3 ; 4 y : 100 ; 140 ; 196 ; 274 ; 384

La calculatrice donne l'ajustement affine $y = 70{,}2x + 78{,}4$, avec $r \approx 0{,}98$.

1. Calcule les quotients de deux valeurs consécutives. Que constates-tu ?

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

2. Propose un modèle mieux adapté et estime y pour $x = 6$ avec chacun des deux modèles.

3. Un coefficient $r = 0{,}98$ garantit-il que le modèle affine est le bon ?

Exercice 18 – Corrélation n’est pas causalité

Dans une ville, on relève chaque mois le nombre de glaces vendues et le nombre de noyades à la piscine municipale. Le coefficient de corrélation entre ces deux variables vaut 0,93.

1. Que peut-on affirmer sur la liaison entre ces deux variables ?

2. Peut-on en conclure que manger des glaces provoque des noyades ?

3. Propose une explication, et dis ce qu’il faudrait faire pour la tester.

Exercice 19 – Quand l’extrapolation devient absurde

Le pourcentage de foyers d’une commune équipés d’une connexion très haut débit est relevé de 2015 à 2020. Avec x = année $-$ 2015 :

x : 0 ; 1 ; 2 ; 3 ; 4 ; 5 y : 62 ; 67 ; 72 ; 77 ; 82 ; 87

La calculatrice donne $y = 5x + 62$ avec $r = 1$.

1. Vérifie l’équation de la droite et justifie que $r = 1$.

2. Que prévoit le modèle pour 2030 ? Commente.

3. À partir de quelle année le modèle devient-il impossible ?

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Exercice 20 – Synthèse : étude d’un chiffre d’affaires

On reprend le chiffre d’affaires (en milliers d’euros) de 2019 à 2024, avec x = année − 2019 :

x : 0 ; 1 ; 2 ; 3 ; 4 ; 5 y : 242 ; 256 ; 279 ; 298 ; 313 ; 337

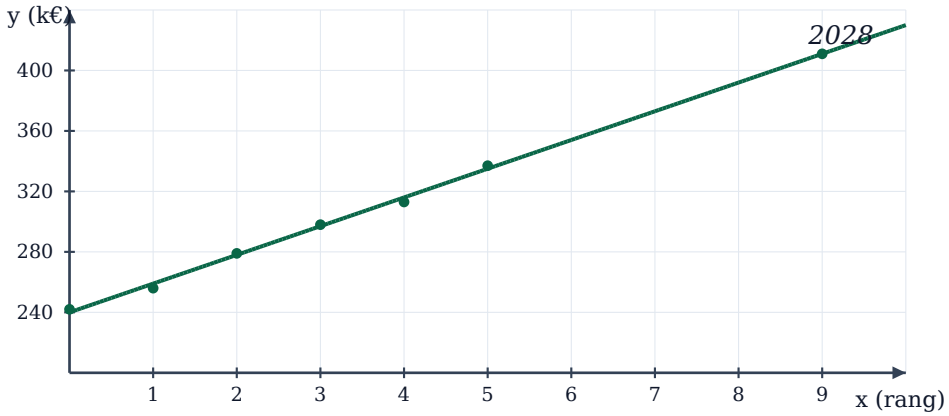
La calculatrice donne $y = 19x + 240$ avec $r \approx 0{,}998$. Le nuage et la droite sont représentés ci-dessous.

1. Interprète les coefficients 19 et 240, ainsi que la valeur de r.

2. Calcule les résidus des années 2019, 2021 et 2024.

3. Estime le chiffre d’affaires de 2026, puis détermine l’année où il dépasserait 400 milliers d’euros.

4. Quelle réserve formuler sur une prévision pour 2050 ?



mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Corrigés détaillés

Les corrections des 20 exercices, dans l'ordre des énoncés.

Correction de l'exercice 1 – Point moyen d'un nuage

Le point moyen G a pour coordonnées $(\overline{x}\,;\overline{y})$: la moyenne des abscisses et la moyenne des ordonnées.

Moyenne des x : $\overline{x} = \dfrac{1+2+3+4+5}{5} = \dfrac{15}{5} = 3$.

Moyenne des y : $\overline{y} = \dfrac{3+5+4+8+10}{5} = \dfrac{30}{5} = 6$.

Le point moyen est donc $G(3\,;6)$.

Interprétation : G est le « centre de gravité » du nuage. Un magasin moyen de cette série emploie 3 vendeurs et réalise 6 milliers d'euros de chiffre d'affaires.

Attention : G n'appartient PAS forcément au nuage. Ici, le point $(3\,;6)$ n'est pas un des cinq relevés — le magasin à 3 vendeurs réalise 4 milliers d'euros, pas 6.

Contrôle de vraisemblance : $\overline{x} = 3$ doit se situer entre le plus petit et le plus grand des x, soit entre 1 et 5. ✔ De même $\overline{y} = 6$ est bien entre 3 et 10. ✔

Ce point servira dans tout le chapitre : la droite d'ajustement des moindres carrés passe TOUJOURS par lui, ce qui fournit une vérification gratuite.

Réponse. $\overline{x} = 3$, $\overline{y} = 6$, donc $G(3\,;6)$.

Rappel de cours. $G(\overline{x}\,;\overline{y})$: on moyenne séparément chaque variable. Le point moyen n'est presque jamais un point du nuage.

Erreur fréquente. Calculer la moyenne des y en la divisant par autre chose que l'effectif. Ici il y a 5 couples, donc on divise par 5 dans les deux cas.

Correction de l'exercice 2 – Décrire un nuage de points

Le nuage monte de la gauche vers la droite : quand x augmente, y augmente globalement aussi.

La corrélation est donc POSITIVE. Plus le magasin a de vendeurs, plus son chiffre d'affaires est élevé — en tendance.

Alignement : les points suivent bien une direction générale, mais ils ne sont pas parfaitement alignés. Le nuage est allongé, pas rectiligne.

Le point qui s'écarte le plus est celui d'abscisse 3 : il est nettement SOUS la droite, avec un chiffre d'affaires de 4 alors que la droite prévoit 6.

Cet écart de 2 milliers d'euros est bien supérieur aux autres, qui sont tous inférieurs à 1. On le retrouvera plus loin sous le nom de RÉSIDU.

Pour la dernière question, on lit sur le graphique : la droite passe par le point d'abscisse 3 à la hauteur 6.

Donc $G(3\,;6)$ appartient bien à la droite tracée.

Ce n'est pas un hasard : la droite d'ajustement des moindres carrés passe toujours par le point moyen du nuage.

Vérification par le calcul, avec l'équation de cette droite $y = 1{,}7x + 0{,}9$: pour $x = 3$, $y = 5{,}1 + 0{,}9 = 6$. ✔

Vocabulaire à retenir : un nuage « allongé » autorise un ajustement affine, un nuage

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

« rond » ou en courbe ne s'y prête pas.

Réponse. Corrélation positive (le nuage monte) · nuage allongé mais pas parfaitement aligné, le point d'abscisse 3 s'écarte le plus (4 au lieu de 6) · $G(3\,;6)$ appartient bien à la droite.

Rappel de cours. La droite des moindres carrés passe toujours par le point moyen G : c'est le contrôle le plus rapide d'une équation d'ajustement.

Erreur fréquente. Conclure « bien corrélé » dès que les points montent. La corrélation a un SENS (positif/négatif) et une FORCE (dispersion autour de la droite) : ce sont deux choses distinctes.

Correction de l'exercice 3 – Moyennes d'une seconde série

Moyenne des dépenses : $\overline{x} = \dfrac{0+2+4+6+8}{5}$.

$0+2+4+6+8 = 20$, donc $\overline{x} = \dfrac{20}{5} = 4$ milliers d'euros.

Moyenne des ventes : $\overline{y} = \dfrac{12+15+19+22+27}{5}$.

On additionne par étapes : $12+15=27$, $27+19=46$, $46+22=68$, $68+27=95$.

Donc $\overline{y} = \dfrac{95}{5} = 19$ milliers d'articles.

Le point moyen est $G(4\,;19)$.

Contrôle de vraisemblance : 4 est bien le milieu de 0 et 8 — normal, les dépenses sont régulièrement espacées. ✔

Pour $\overline{y}$, le contrôle est moins immédiat : 19 doit se situer entre 12 et 27, et plutôt vers le milieu. ✔

Remarque utile : la moyenne des y n'est pas la valeur centrale de la liste. Ici la valeur centrale est 19 et la moyenne aussi, mais c'est une coïncidence due à la régularité de la série.

Réponse. $\overline{x} = 4$, $\overline{y} = 19$, donc $G(4\,;19)$.

Rappel de cours. On additionne toutes les valeurs d'une variable et on divise par leur NOMBRE. Une erreur d'addition se repère en vérifiant que la moyenne tombe entre le minimum et le maximum.

Erreur fréquente. Oublier la valeur 0 dans le calcul de $\overline{x}$ et diviser par 4. Un relevé nul reste un relevé : l'effectif est bien 5.

Correction de l'exercice 4 – Utiliser une droite d'ajustement

Pour vérifier le passage par G, on remplace x par 4 dans l'équation.

$y = 1{,}85 \times 4 + 11{,}6 = 7{,}4 + 11{,}6 = 19$.

On obtient exactement $\overline{y} = 19$, donc la droite passe bien par $G(4\,;19)$. ✔ C'est la vérification standard d'un ajustement.

Estimation pour $x = 10$: $y = 1{,}85 \times 10 + 11{,}6 = 18{,}5 + 11{,}6 = 30{,}1$.

On peut donc estimer les ventes à environ 30 100 articles pour une dépense publicitaire de 10 000 €.

Le mot « estimer » est important : la droite est un modèle, pas une mesure. La valeur réelle sera proche, pas identique.

Interprétation du coefficient directeur 1,85 : quand la dépense augmente de 1 millier d'euros, les ventes augmentent d'environ 1,85 millier d'articles.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Autrement dit : 1 000 € de publicité supplémentaires rapportent environ 1 850 articles vendus de plus, selon ce modèle.

Le coefficient 11,6 s'interprète aussi : c'est l'estimation des ventes sans aucune publicité, ce que confirme le relevé réel de 12 pour $x = 0$. ✔

Contrôle de cohérence de la prévision : 30,1 prolonge naturellement la série $12, 15, 19, 22, 27$, qui progresse d'environ 3,7 tous les 2 milliers d'euros. ✔

Réponse. $1{,}85 \times 4 + 11{,}6 = 19$, la droite passe bien par G · ventes estimées à 30,1 milliers d'articles pour 10 k€ · le coefficient 1,85 signifie +1 850 articles par millier d'euros investi.

Rappel de cours. Le coefficient directeur d'un ajustement se lit toujours comme un TAUX : « quand x augmente de 1, y augmente de a ».

Erreur fréquente. Interpréter 1,85 comme un nombre d'articles. C'est une variation par unité de x : 1,85 millier d'articles PAR millier d'euros.

Correction de l'exercice 5 – Interpolation ou extrapolation

Pour $x = 5$: $y = 1{,}85 \times 5 + 11{,}6 = 9{,}25 + 11{,}6 = 20{,}85$.

Pour $x = 12$: $y = 1{,}85 \times 12 + 11{,}6 = 22{,}2 + 11{,}6 = 33{,}8$.

Pour $x = 40$: $y = 1{,}85 \times 40 + 11{,}6 = 74 + 11{,}6 = 85{,}6$.

Vocabulaire : on parle d'INTERPOLATION quand la valeur de x est comprise dans l'intervalle des données observées, et d'EXTRAPOLATION quand elle est en dehors.

Ici les données vont de 0 à 8. Donc $x = 5$ est une interpolation, tandis que $x = 12$ et $x = 40$ sont des extrapolations.

La moins fiable est de loin l'estimation pour $x = 40$, qui sort très largement du domaine observé.

Pourquoi ? Parce que rien ne garantit que la relation reste affine au-delà des données. Un budget publicitaire cinq fois plus grand que tout ce qui a été testé peut saturer le marché, ou rencontrer une limite de production.

Le modèle prévoit 85 600 articles vendus, ce qui n'est étayé par aucune observation.

Règle pratique : une extrapolation proche du domaine (comme $x = 12$) reste raisonnable ; une extrapolation lointaine doit toujours être assortie d'une réserve explicite.

À retenir : un ajustement décrit ce qu'on a observé, il ne prouve rien sur ce qu'on n'a pas observé.

Réponse. 20,85 ; 33,8 ; 85,6 · $x = 5$: interpolation · $x = 12$ et $x = 40$: extrapolations · celle en $x = 40$ est la moins fiable, car très éloignée des données (0 à 8).

Rappel de cours. Interpolation = à l'intérieur du domaine étudié (fiable) · extrapolation = à l'extérieur (d'autant moins fiable qu'on s'éloigne).

Erreur fréquente. Traiter toutes les prévisions sur le même plan. Le calcul est identique, mais la CONFIANCE qu'on peut leur accorder ne l'est pas du tout.

Correction de l'exercice 6 – Ajustement par les points extrêmes

Les deux points extrêmes sont $A(1\,;3)$ et $B(5\,;10)$.

Coefficient directeur : $a = \dfrac{y_B - y_A}{x_B - x_A} = \dfrac{10-3}{5-1} = \dfrac{7}{4} = 1{,}75$.

Ordonnée à l'origine : on utilise A, donc $3 = 1{,}75 \times 1 + b$, d'où $b = 3 - 1{,}75 = 1{,}25$.

Équation : $y = 1{,}75x + 1{,}25$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Vérification avec le point B : $1{,}75 \times 5 + 1{,}25 = 8{,}75 + 1{,}25 = 10$. ✔

Passage par G : pour $x = 3$, la droite donne $y = 5{,}25 + 1{,}25 = 6{,}5$.

Or $\overline{y} = 6$. La droite ne passe donc PAS par le point moyen : elle passe 0,5 au-dessus.

Critique de la méthode : elle est très rapide, mais elle ne tient compte que de DEUX points sur cinq. Les trois points intermédiaires n'interviennent pas du tout.

Elle est donc très sensible aux valeurs extrêmes : si le magasin à 5 vendeurs avait réalisé 6 au lieu de 10, toute la droite basculerait, alors que le reste du nuage n'aurait pas bougé.

C'est précisément pour cela qu'on lui préfère la méthode des moindres carrés, qui utilise tous les points et passe par G.

Réponse. $y = 1{,}75x + 1{,}25$ · elle ne passe pas par G (elle donne 6,5 au lieu de 6) · méthode rapide mais fragile, car seuls deux points sur cinq sont utilisés.

Rappel de cours. Droite par deux points : $a = \dfrac{y_B - y_A}{x_B - x_A}$, puis b en remplaçant les coordonnées de l'un des deux.

Erreur fréquente. Calculer $a = \dfrac{x_B - x_A}{y_B - y_A}$, soit l'inverse. Le coefficient directeur est la variation de y DIVISÉE par celle de x.

Correction de l'exercice 7 – Droite des moindres carrés

Passage par G : $1{,}7 \times 3 + 0{,}9 = 5{,}1 + 0{,}9 = 6$, soit exactement $\overline{y}$. ✔

C'est la propriété caractéristique de la droite des moindres carrés : elle passe toujours par le point moyen, contrairement à la droite des points extrêmes.

Comparaison pour $x = 2$: moindres carrés $1{,}7 \times 2 + 0{,}9 = 4{,}3$; points extrêmes $1{,}75 \times 2 + 1{,}25 = 4{,}75$.

La valeur réelle est 5. L'ajustement par les points extrêmes est ici plus proche, mais ce n'est pas une règle.

Comparaison pour $x = 5$: moindres carrés $8{,}5 + 0{,}9 = 9{,}4$; points extrêmes $8{,}75 + 1{,}25 = 10$.

La droite des points extrêmes passe exactement par ce point — forcément, puisqu'elle a été construite à partir de lui. Cela ne prouve pas qu'elle est meilleure.

Ce que fait la droite des moindres carrés : elle rend la somme des CARRÉS des écarts verticaux la plus petite possible, en tenant compte de tous les points à la fois.

Estimation pour 7 vendeurs : $y = 1{,}7 \times 7 + 0{,}9 = 11{,}9 + 0{,}9 = 12{,}8$ milliers d'euros.

Il s'agit d'une extrapolation (les données s'arrêtent à 5 vendeurs), mais elle reste proche du domaine observé : elle est donc plausible, à condition de le préciser.

Contrôle de vraisemblance : 12,8 prolonge la progression du nuage, qui gagne environ 1,7 par vendeur supplémentaire. ✔

Réponse. $1{,}7 \times 3 + 0{,}9 = 6$, la droite passe par G · en $x = 2$: 4,3 contre 4,75 ; en $x = 5$: 9,4 contre 10 · pour 7 vendeurs, on estime 12,8 milliers d'euros.

Rappel de cours. La droite des moindres carrés minimise la somme des carrés des écarts verticaux et passe par G. En Première, on l'obtient à la calculatrice.

Erreur fréquente. Juger un ajustement sur un seul point. Une droite qui passe exactement par un point donné peut être globalement moins bonne que l'autre.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l’exercice 8 – Calculer des résidus

Le résidu mesure l’écart VERTICAL entre le point réel et la droite : positif si le point est au-dessus, négatif s’il est en dessous.

Pour $x = 1$: estimation $1{,}7 + 0{,}9 = 2{,}6$, valeur réelle 3, résidu $3 - 2{,}6 = +0{,}4$.

Pour $x = 2$: estimation $3{,}4 + 0{,}9 = 4{,}3$, valeur réelle 5, résidu $+0{,}7$.

Pour $x = 3$: estimation $5{,}1 + 0{,}9 = 6$, valeur réelle 4, résidu $4 - 6 = -2$.

Pour $x = 4$: estimation $6{,}8 + 0{,}9 = 7{,}7$, valeur réelle 8, résidu $+0{,}3$.

Pour $x = 5$: estimation $8{,}5 + 0{,}9 = 9{,}4$, valeur réelle 10, résidu $+0{,}6$.

Somme des résidus : $0{,}4 + 0{,}7 - 2 + 0{,}3 + 0{,}6$.

On regroupe les positifs : $0{,}4 + 0{,}7 + 0{,}3 + 0{,}6 = 2$. Puis $2 - 2 = 0$.

La somme des résidus est donc EXACTEMENT nulle. Ce n’est pas une coïncidence : c’est une propriété de la droite des moindres carrés, qui répartit les écarts de part et d’autre.

C’est d’ailleurs pour cette raison qu’on minimise la somme des CARRÉS des résidus : la somme des résidus simples serait nulle pour de nombreuses droites, elle ne mesurerait rien.

Lecture du tableau : quatre résidus sont petits (moins de 0,7), un seul est grand (-2). Le magasin à 3 vendeurs sous-performe nettement par rapport au modèle.

C’est exactement le point repéré graphiquement au début du chapitre. ✔ Les deux approches concordent.

Réponse. Résidus : $+0{,}4$; $+0{,}7$; -2 ; $+0{,}3$; $+0{,}6$. Leur somme vaut 0, propriété de la droite des moindres carrés.

Rappel de cours. Résidu = valeur réelle − valeur estimée. Un gros résidu isolé signale un individu atypique, pas forcément une erreur.

Erreur fréquente. Calculer « estimée moins réelle ». Le signe s’inverse et l’interprétation aussi : un résidu positif doit signifier que le point est AU-DESSUS de la droite.

Correction de l’exercice 9 – Interpréter un coefficient de corrélation

Le coefficient de corrélation linéaire r est toujours compris entre -1 et 1.

Son SIGNE donne le sens de la liaison : positif si les deux variables augmentent ensemble, négatif si l’une augmente quand l’autre diminue.

Sa VALEUR ABSOLUE donne la force : proche de 1, les points sont presque alignés ; proche de 0, le nuage n’a pas de direction privilégiée.

$r = 0{,}996$: liaison positive et très forte. Les points sont quasiment alignés, l’ajustement affine est tout à fait pertinent.

$r = 0{,}92$: liaison positive et forte, mais avec une dispersion visible. L’ajustement reste pertinent, tout en donnant des prévisions moins précises.

$r = -0{,}85$: liaison NÉGATIVE et forte. Quand le prix augmente, la quantité vendue diminue — ce qui est économiquement attendu.

Le signe négatif se traduit par un coefficient directeur négatif dans l’équation d’ajustement : la droite descend.

$r = 0{,}15$: liaison quasi inexistante. Le nuage est « rond », sans direction. Un ajustement affine n’aurait aucun sens ici.

Ce dernier exemple est instructif : entre pointure et note en mathématiques, on n’attend

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

évidemment aucun lien, et le coefficient le confirme.

Mise en garde essentielle : r ne mesure QUE la liaison linéaire. Un nuage parfaitement disposé sur une parabole peut donner un r faible alors que la liaison est parfaite — elle n'est simplement pas droite.

Réponse. 0,996 : positive très forte, ajustement pertinent · 0,92 : positive forte, ajustement pertinent · −0,85 : négative forte, droite décroissante · 0,15 : liaison quasi nulle, aucun ajustement affine.

Rappel de cours. $-1 \leq r \leq 1$: le signe donne le sens, la valeur absolue la force. r ne détecte que les liaisons LINÉAIRES.

Erreur fréquente. Comparer −0,85 et 0,15 en disant que le second est « plus grand donc mieux corrélé ». On compare les valeurs absolues : $0{,}85 > 0{,}15$.

Correction de l'exercice 10 – Changer l'origine des années

Les valeurs de x sont 0, 1, 2, 3, 4, 5 : ce sont les rangs des années, à partir de 2019.

Intérêt : les calculs portent sur de petits nombres au lieu de nombres à quatre chiffres, ce qui limite les erreurs et rend les coefficients lisibles.

Second intérêt, plus important : l'ordonnée à l'origine devient interprétable. Avec x = année, le « b » correspondrait à l'an 0, ce qui n'a aucun sens.

Moyenne des rangs : $\overline{x} = \dfrac{0+1+2+3+4+5}{6} = \dfrac{15}{6} = 2{,}5$.

Moyenne des chiffres d'affaires : 242 + 256 = 498, +279 = 777, +298 = 1 075, +313 = 1 388, +337 = 1 725.

Donc $\overline{y} = \dfrac{1\,725}{6} = 287{,}5$ milliers d'euros.

Le point moyen est $G(2{,}5\,;\,287{,}5)$.

Vérification du passage par G : $19 \times 2{,}5 + 240 = 47{,}5 + 240 = 287{,}5$. ✔ C'est exactement $\overline{y}$.

Interprétation des coefficients : le chiffre d'affaires augmente d'environ 19 milliers d'euros par an, et valait 240 milliers d'euros au rang 0, c'est-à-dire en 2019.

Contrôle : la valeur réelle en 2019 est 242, très proche des 240 du modèle. ✔ Le changement d'origine rend ce contrôle immédiat.

Réponse. $x = 0$ à 5 · le changement d'origine simplifie les calculs et rend b interprétable · $\overline{x} = 2{,}5$, $\overline{y} = 287{,}5$ · $19 \times 2{,}5 + 240 = 287{,}5$, la droite passe bien par G.

Rappel de cours. Avec des années, on pose toujours x = année − année de départ. Il faut ensuite penser à reconvertir les rangs en années dans la conclusion.

Erreur fréquente. Garder $x = 2019, 2020, \ldots$ et annoncer une ordonnée à l'origine « en l'an 0 ». Le coefficient b n'aurait alors aucune signification concrète.

Correction de l'exercice 11 – Méthode de Mayer

Le nuage compte 6 points : on le partage en deux groupes de 3, en respectant l'ordre des abscisses.

Premier sous-nuage : les points d'abscisses 1, 2, 3.

Son point moyen : $x_1 = \dfrac{1+2+3}{3} = 2$ et $y_1 = \dfrac{3+5+7}{3} = \dfrac{15}{3} = 5$. Donc $G_1(2\,;\,5)$.

Second sous-nuage : les points d'abscisses 4, 5, 6.

Son point moyen : $x_2 = \dfrac{4+5+6}{3} = 5$ et $y_2 = \dfrac{10+11+12}{3} = \dfrac{33}{3} = 11$. Donc $G_2(5\,;\,11)$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Coefficient directeur de la droite (G_1G_2) : $a = \dfrac{11-5}{5-2} = \dfrac{6}{3} = 2$.

Ordonnée à l'origine, avec G_1 : $5 = 2 \times 2 + b$, donc $b = 5 - 4 = 1$.

Équation de la droite de Mayer : $y = 2x + 1$.

Vérification avec G_2 : $2 \times 5 + 1 = 11$. ✔

Contrôle supplémentaire, très utile : cette droite doit passer par le point moyen GLOBAL du nuage.

$\overline{x} = \dfrac{1+2+3+4+5+6}{6} = 3{,}5$ et $\overline{y} = \dfrac{3+5+7+10+11+12}{6} = \dfrac{48}{6} = 8$.

Or $2 \times 3{,}5 + 1 = 8$. ✔ La droite de Mayer passe bien par $G(3{,}5\,;8)$ — c'est toujours le cas, puisque G est le milieu de G_1G_2 quand les deux groupes ont le même effectif.

Avantage de la méthode par rapport aux points extrêmes : elle utilise TOUS les points du nuage, et se calcule entièrement à la main.

Réponse. $G_1(2\,;5)$ et $G_2(5\,;11)$, donc la droite de Mayer a pour équation $y = 2x + 1$; elle passe bien par le point moyen $G(3{,}5\,;8)$.

Rappel de cours. Méthode de Mayer : deux sous-nuages de même effectif, leurs points moyens, la droite qui les joint. Elle passe toujours par G.

Erreur fréquente. Partager le nuage n'importe comment. On sépare selon les abscisses CROISSANTES, en deux groupes de même effectif (ou presque, si l'effectif est impair).

Correction de l'exercice 12 – Comparer deux ajustements

Pour $x = 3$, Mayer donne $2 \times 3 + 1 = 7$.

Les moindres carrés donnent $1{,}886 \times 3 + 1{,}4 = 5{,}658 + 1{,}4 = 7{,}058$, soit environ 7,06.

L'écart est de 0,06 : totalement négligeable. La valeur réelle est d'ailleurs exactement 7.

Pour $x = 6$, Mayer donne 13 et les moindres carrés
$1{,}886 \times 6 + 1{,}4 = 11{,}316 + 1{,}4 = 12{,}72$.

L'écart atteint 0,28, toujours faible. La valeur réelle est 12, donc les moindres carrés sont ici un peu plus proches.

Pour $x = 10$, Mayer donne 21 et les moindres carrés $18{,}86 + 1{,}4 = 20{,}26$.

L'écart monte à 0,74 : il grandit à mesure qu'on s'éloigne du nuage.

Conclusion : les deux méthodes donnent des résultats très proches à l'intérieur du domaine des données, et divergent progressivement en extrapolation.

C'est un phénomène général : deux droites d'ajustement différentes se croisent au voisinage du point moyen et s'écartent de part et d'autre, comme deux droites sécantes.

Conséquence pratique : le choix de la méthode importe peu pour interpoler, mais devient sensible dès qu'on extrapole loin.

Vérification que les deux droites passent bien par $G(3{,}5\,;8)$: Mayer donne 8 ✔, et les moindres carrés $1{,}886 \times 3{,}5 + 1{,}4 = 6{,}601 + 1{,}4 = 8{,}001$, soit 8 à l'arrondi près. ✔

Cette coïncidence n'en est pas une : les deux méthodes passent par le point moyen, donc les droites se COUPENT en G. C'est là que l'écart entre elles est nul.

Réponse. En $x = 3$: 7 contre 7,06 · en $x = 6$: 13 contre 12,72 · en $x = 10$: 21 contre 20,26 · les deux méthodes concordent près des données et divergent en extrapolation, car elles se coupent en G.

Rappel de cours. Deux ajustements passant par G se coupent en G : leur écart est nul au centre du nuage et croît quand on s'en éloigne.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Erreur fréquente. Choisir un modèle parce qu'il donne « un plus grand résultat ». Le critère est la qualité de l'ajustement, jamais la valeur obtenue.

Correction de l'exercice 13 – Prévision et seuil

Estimation au rang 8 : $y = 2 \times 8 + 1 = 17$ centaines de commandes, soit 1 700 commandes.

Pour le seuil, on résout l'inéquation $2x + 1 > 25$.

$2x > 24$, donc $x > 12$.

Comme le rang est un nombre entier de mois, le premier rang qui convient est $x = 13$.

Vérification : au rang 12, $y = 25$ exactement — ce n'est pas « plus de 25 ». Au rang 13, $y = 27 > 25$. ✔

L'inégalité étant STRICTE, le rang 12 doit être écarté. Avec « au moins 25 », la réponse aurait été 12.

Pour la date : si le rang 1 est janvier 2026, alors le rang 13 est douze mois plus tard.

Douze mois après janvier 2026, on est en janvier 2027.

Attention au décalage : le rang 13 n'est pas « treize mois après janvier 2026 », mais le treizième mois de la série, donc janvier 2027.

Contrôle par énumération partielle : rang 1 janvier 2026, rang 12 décembre 2026, rang 13 janvier 2027. ✔

Réserve à formuler : le rang 13 sort du domaine des données (six mois observés). C'est une extrapolation, et rien ne garantit que la croissance restera régulière pendant plus d'un an.

Réponse. 1 700 commandes au rang 8 · le seuil est dépassé à partir du rang 13 · soit janvier 2027, avec la réserve que cette prévision est une extrapolation lointaine.

Rappel de cours. Après avoir résolu l'inéquation, on revient toujours à l'unité concrète (rang entier, puis date) — c'est la question posée.

Erreur fréquente. Répondre « rang 12 » à une inégalité stricte. Au rang 12 le modèle donne exactement 25, ce qui ne dépasse pas 25.

Correction de l'exercice 14 – Linéariser par changement de variable

Un nuage courbé ne se prête pas à un ajustement affine. L'astuce consiste à transformer la variable y pour redresser le nuage.

Calcul des $z = \frac{1}{y}$: pour $y = 2$, $z = 0{,}5$; pour $y = 1$, $z = 1$.

Pour $y = 0{,}667$, $z = \frac{1}{0{,}667} \approx 1{,}5$; pour $y = 0{,}5$, $z = 2$.

La série des z est donc 0,5 ; 1 ; 1,5 ; 2.

Ces valeurs augmentent de 0,5 à chaque fois que x augmente de 1 : les points $(x\,;z)$ sont PARFAITEMENT alignés.

Relation entre z et x : le coefficient directeur vaut 0,5, et pour $x = 1$ on a $z = 0{,}5$, donc $b = 0{,}5 - 0{,}5 = 0$.

Ainsi $z = 0{,}5x$, une relation de proportionnalité.

Retour à y : puisque $z = \frac{1}{y}$, on a $y = \frac{1}{z} = \frac{1}{0{,}5x} = \frac{2}{x}$.

Vérification sur les données : $\frac{2}{1} = 2$ ✔, $\frac{2}{2} = 1$ ✔, $\frac{2}{3} \approx 0{,}667$ ✔, $\frac{2}{4} = 0{,}5$ ✔.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Estimation pour $x = 10$: $z = 0{,}5 \times 10 = 5$, donc $y = \frac{1}{5} = 0{,}2$.

Vérification directe avec la formule : $y = \frac{2}{10} = 0{,}2$. ✔ Les deux chemins concordent.

Méthode générale à retenir : quand un nuage est courbé, on cherche une transformation ($\frac{1}{y}$, y^2, $\sqrt{y}$…) qui l'aligne, on ajuste dans le nouveau repère, puis on REVIENT à la variable d'origine.

Réponse. z : 0,5 ; 1 ; 1,5 ; 2, parfaitement alignés · $z = 0{,}5x$, donc $y = \frac{2}{x}$ · pour $x = 10$, $y = 0{,}2$.

Rappel de cours. Linéarisation : on ajuste $z = f(y)$ en fonction de x, puis on revient à y. Ne jamais oublier cette dernière étape.

Erreur fréquente. Donner $z = 5$ comme réponse finale. La question porte sur y, il faut repasser par $y = \frac{1}{z}$.

Correction de l'exercice 15 – Lecture graphique d'un ajustement

Ordonnée à l'origine : on lit l'ordonnée du point où la droite coupe l'axe vertical, soit 1.

Coefficient directeur : on part d'un point de la droite et on avance de 1 vers la droite. On constate qu'il faut monter de 2.

Le coefficient directeur vaut donc 2, et l'équation lue est $y = 2x + 1$, conforme à la droite de Mayer calculée précédemment. ✔

Estimation graphique pour $x = 7$: on prolonge mentalement la droite jusqu'à l'abscisse 7, et on lit environ 15.

Vérification par le calcul : $y = 2 \times 7 + 1 = 15$. ✔ La lecture graphique est exacte ici, parce que les valeurs sont entières.

Pour identifier les points au-dessus de la droite, on compare chaque valeur réelle à la valeur estimée.

$x = 1$: réel 3, droite 3 — le point est SUR la droite.

$x = 2$: réel 5, droite 5 — également sur la droite.

$x = 3$: réel 7, droite 7 — sur la droite aussi.

$x = 4$: réel 10, droite 9 — le point est AU-DESSUS, d'une unité.

$x = 5$: réel 11, droite 11 — sur la droite. Et $x = 6$: réel 12, droite 13 — le point est EN DESSOUS.

Un seul point est donc strictement au-dessus ($x = 4$) et un seul strictement en dessous ($x = 6$) ; les autres sont exactement sur la droite.

Contrôle : la somme des écarts vaut $0 + 0 + 0 + 1 + 0 - 1 = 0$. ✔ L'ajustement est équilibré, ce qui est bon signe.

Réponse. Ordonnée à l'origine 1, coefficient directeur 2, soit $y = 2x + 1$ · pour $x = 7$, on lit et on calcule 15 · seul le point d'abscisse 4 est au-dessus de la droite (celui d'abscisse 6 est en dessous).

Rappel de cours. Lire un coefficient directeur : on avance de 1 sur l'axe horizontal et on mesure la montée. Lire b : on regarde l'intersection avec l'axe vertical.

Erreur fréquente. Lire le coefficient directeur sur deux points du NUAGE au lieu de deux points de la droite. Les points du nuage ne sont pas sur la droite en général.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l’exercice 16 – Effet d’un point aberrant

Pour vérifier l’alignement, on regarde les écarts successifs des y : $8-5=3$, $11-8=3$, $14-11=3$, $17-14=3$, $20-17=3$.

Les x augmentent de 1 et les y de 3 à chaque fois : les points sont parfaitement alignés sur une droite de coefficient directeur 3.

Vérification de l’équation : pour $x=1$, $3\times1+2=5$ ✔ ; pour $x=6$, $3\times6+2=20$ ✔.

Un alignement parfait donne $r=1$, la valeur maximale possible. ✔

Avec la faute de frappe, le coefficient directeur s’effondre de 3 à 0,43, soit sept fois moins.

Le coefficient de corrélation chute de 1 à 0,14 : la liaison, parfaite, devient statistiquement inexistante.

Un SEUL point sur six a suffi à détruire l’ajustement. C’est la principale fragilité de la méthode des moindres carrés : elle est très sensible aux valeurs extrêmes.

Pourquoi ? Parce qu’elle minimise la somme des CARRÉS des écarts : un écart de 18 pèse 324, alors que dix écarts de 1 ne pèsent que 10. Le point aberrant domine tout le calcul.

Que faire en pratique ? D’abord TRACER le nuage : un point aberrant se voit immédiatement, alors qu’il est invisible dans un tableau de nombres.

Ensuite, chercher l’origine de l’écart : erreur de saisie (comme ici), erreur de mesure, ou individu réellement atypique.

On ne supprime jamais un point parce qu’il dérange. On l’écarte s’il est établi qu’il s’agit d’une erreur ; sinon on le conserve, en le signalant et en discutant son influence.

Contrôle final : après correction de la saisie, on doit retrouver $r=1$ et $y=3x+2$. ✔

Réponse. Écarts constants de 3 : la série est parfaitement alignée, $y=3x+2$ et $r=1$ · avec l’erreur, le coefficient tombe à 0,43 et r à 0,14 · il faut tracer le nuage, identifier l’origine du point aberrant, et ne l’écarter que s’il s’agit d’une erreur avérée.

Rappel de cours. Toujours tracer le nuage AVANT d’exploiter un ajustement : un point aberrant fausse à la fois la droite et le coefficient de corrélation.

Erreur fréquente. Faire confiance à un r faible pour conclure « pas de lien ». Ici le lien est parfait ; c’est une seule donnée erronée qui masque tout.

Correction de l’exercice 17 – Corrélation forte, modèle inadapté

Quotients successifs : $\frac{140}{100}=1{,}4$; $\frac{196}{140}=1{,}4$.

Puis $\frac{274}{196}\approx1{,}398$ et $\frac{384}{274}\approx1{,}402$.

Les quotients sont donc tous égaux à 1,4 à l’arrondi près : ce n’est pas une croissance affine, mais une croissance GÉOMÉTRIQUE.

Autrement dit, le nombre d’abonnés augmente d’environ 40 % par an — on retrouve les chapitres 2 et 3.

Modèle adapté : $y=100\times1{,}4^{x}$, une suite géométrique de premier terme 100 et de raison 1,4.

Vérification du modèle sur les données : $100\times1{,}4^{2}=196$ ✔ et $100\times1{,}4^{4}=384{,}16\approx384$. ✔

Estimation pour $x=6$ avec le modèle affine : $70{,}2\times6+78{,}4=421{,}2+78{,}4=499{,}6$, soit environ 500 milliers d’abonnés.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Estimation avec le modèle géométrique : $100 \times 1{,}4^{6} \approx 752{,}95$, soit environ 753 milliers d'abonnés.

L'écart est de 253 milliers d'abonnés, soit plus de 50 % d'erreur relative. Le choix du modèle n'est donc pas un détail.

Pour la dernière question : NON, un coefficient $r = 0{,}98$ ne garantit rien.

Sur un intervalle court, une courbe exponentielle ressemble beaucoup à une droite : le coefficient de corrélation reste élevé alors que la nature du phénomène est tout autre.

La bonne démarche est de regarder la FORME du nuage et les quotients successifs, pas seulement r. Ici, les quotients constants trahissent immédiatement le modèle géométrique.

Contrôle par les résidus du modèle affine : en $x = 0$, il prédit 78,4 pour une valeur réelle de 100 ; en $x = 2$, 218,8 pour 196 ; en $x = 4$, 359,2 pour 384. Les résidus sont d'abord positifs, puis négatifs, puis positifs — signe caractéristique d'un modèle mal choisi.

Réponse. Quotients tous égaux à 1,4 : croissance géométrique de +40 % par an · modèle $y = 100 \times 1{,}4^{x}$ · en $x = 6$: 499,6 (affine) contre 753 (géométrique) · un r élevé ne garantit pas que le modèle affine soit le bon.

Rappel de cours. Écarts constants → modèle affine · quotients constants → modèle géométrique. On teste TOUJOURS les deux avant de choisir.

Erreur fréquente. Se fier au seul coefficient de corrélation. Une exponentielle sur un intervalle court donne facilement un r supérieur à 0,95.

Correction de l'exercice 18 – Corrélation n'est pas causalité

Un coefficient de 0,93 indique une liaison linéaire positive et FORTE entre les deux variables.

Statistiquement, cela signifie que les mois où l'on vend beaucoup de glaces sont aussi ceux où l'on compte davantage de noyades.

C'est tout ce que l'on peut affirmer : une liaison statistique, constatée sur les données.

En revanche, on ne peut ABSOLUMENT PAS en conclure que les glaces provoquent des noyades. La corrélation ne dit rien du mécanisme.

Une corrélation forte peut avoir trois origines : une vraie causalité, une causalité inverse, ou une troisième variable qui influence les deux.

Ici, la troisième explication s'impose : la TEMPÉRATURE. Quand il fait chaud, on achète plus de glaces ET on va plus souvent se baigner, donc le risque de noyade augmente.

Cette troisième variable s'appelle un facteur commun, ou variable de confusion. Elle crée une corrélation entre deux grandeurs qui n'ont aucun lien direct.

Comment tester cette explication ? En étudiant la corrélation entre glaces et noyades à température comparable, par exemple en ne gardant que les mois où il a fait entre 25 et 30 °C.

Si la corrélation s'effondre à température constante, c'est bien la chaleur qui expliquait tout.

On peut aussi calculer directement la corrélation entre température et chacune des deux variables : on s'attend à trouver deux coefficients élevés.

Autre test possible : comparer deux villes de climats différents, ou une même ville sur des saisons contrastées.

Règle à retenir pour tout le chapitre : une corrélation, même de 0,99, n'établit jamais une causalité. Elle signale une piste, qu'il faut ensuite expliquer par un mécanisme.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Réponse. On peut seulement affirmer une liaison statistique positive et forte · non, aucune causalité ne s'en déduit · la température est un facteur commun, qu'on peut tester en comparant les mois à température comparable.

Rappel de cours. Corrélation ≠ causalité. Trois lectures possibles : x cause y, y cause x, ou une troisième variable cause les deux.

Erreur fréquente. Transformer une corrélation forte en explication. Le coefficient mesure une coïncidence de variations, pas un mécanisme.

Correction de l'exercice 19 – Quand l'extrapolation devient absurde

Les y augmentent de 5 à chaque fois que x augmente de 1 : $67 - 62 = 5$, $72 - 67 = 5$, et ainsi de suite.

Les points sont donc parfaitement alignés, sur une droite de coefficient directeur 5.

Ordonnée à l'origine : pour $x = 0$, $y = 62$, donc $b = 62$. L'équation $y = 5x + 62$ est confirmée. ✔

Un alignement parfait donne $r = 1$: c'est le cas idéal, où le modèle affine décrit exactement les données observées.

Prévision pour 2030 : le rang est $2030 - 2015 = 15$.

$y = 5 \times 15 + 62 = 75 + 62 = 137$.

Le modèle prévoit donc 137 % de foyers équipés — ce qui est IMPOSSIBLE, un pourcentage de foyers ne pouvant dépasser 100 %.

Cet exemple est instructif : l'ajustement est parfait sur les données ($r = 1$), et pourtant la prévision est absurde. La qualité de l'ajustement ne dit rien sur sa validité hors du domaine observé.

Pour la dernière question, on cherche quand le modèle atteint 100 % : $5x + 62 = 100$ donne $5x = 38$, donc $x = 7{,}6$.

Cela correspond au courant de l'année 2015 + 7,6, soit 2022. Dès 2023 ($x = 8$), le modèle annonce 102 % : il est déjà invalide.

Le modèle affine n'est donc utilisable que sur quelques années après les données, et pas au-delà de 2022.

Explication du phénomène : un taux d'équipement SATURE. La croissance ralentit à mesure qu'on approche de 100 %, car il reste de moins en moins de foyers à équiper.

Un modèle réaliste devrait plafonner, comme la courbe de refroidissement du chapitre 4 qui tendait vers la température ambiante sans jamais la dépasser.

Conclusion générale : avant d'extrapoler, il faut toujours se demander si la grandeur étudiée a une LIMITE physique — pourcentage, capacité, population maximale.

Réponse. Écarts constants de 5 et $y = 5x + 62$, donc alignement parfait et $r = 1$ · pour 2030 le modèle prévoit 137 %, ce qui est impossible · il devient invalide dès qu'il dépasse 100 %, soit courant 2022 ($x = 7{,}6$).

Rappel de cours. $r = 1$ signifie que le modèle colle aux DONNÉES OBSERVÉES. Il ne dit rien de sa validité en dehors du domaine étudié.

Erreur fréquente. Annoncer 137 % sans commentaire. Une prévision doit toujours être confrontée au bon sens et aux limites physiques de la grandeur.

Correction de l'exercice 20 – Synthèse : étude d'un chiffre d'affaires

Le coefficient directeur 19 signifie que le chiffre d'affaires augmente en moyenne de 19 milliers d'euros par an sur la période étudiée.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

L'ordonnée à l'origine 240 correspond au rang 0, c'est-à-dire à l'année 2019 : le modèle y estime 240 milliers d'euros, pour une valeur réelle de 242.

Le coefficient $r \approx 0{,}998$ est très proche de 1 : la liaison est positive et quasi parfaitement linéaire. L'ajustement affine est donc pleinement justifié.

Résidu de 2019 ($x = 0$) : estimation 240, valeur réelle 242, donc résidu +2.

Résidu de 2021 ($x = 2$) : estimation $19 \times 2 + 240 = 278$, valeur réelle 279, donc résidu +1.

Résidu de 2024 ($x = 5$) : estimation $19 \times 5 + 240 = 335$, valeur réelle 337, donc résidu +2.

Ces résidus sont tous petits devant des valeurs de l'ordre de 300 : moins de 1 % d'écart, ce qui confirme la qualité de l'ajustement. ✔

Estimation pour 2026 : le rang vaut $2026 - 2019 = 7$, donc $y = 19 \times 7 + 240 = 133 + 240 = 373$ milliers d'euros.

Pour le seuil, on résout $19x + 240 > 400$, soit $19x > 160$, donc $x > \dfrac{160}{19} \approx 8{,}42$.

Le premier rang entier qui convient est $x = 9$, ce qui correspond à l'année $2019 + 9 = 2028$.

Vérification : au rang 8 (2027), le modèle donne $392 < 400$; au rang 9, il donne $411 > 400$. ✔

Pour 2050, le rang serait 31, et le modèle prévoirait $19 \times 31 + 240 = 589 + 240 = 829$ milliers d'euros.

Réserve à formuler : il s'agit d'une extrapolation à 26 ans au-delà des données, alors que l'ajustement repose sur six années seulement.

Rien ne garantit que la croissance restera linéaire aussi longtemps : le marché peut saturer, la concurrence évoluer, l'entreprise changer d'activité.

Contrôle de cohérence de l'ensemble : la droite passe par $G(2{,}5\,;\,287{,}5)$, puisque $19 \times 2{,}5 + 240 = 287{,}5$. ✔ Tous les résultats de l'exercice s'appuient sur un ajustement validé.

Réponse. +19 k€ par an, 240 k€ estimés en 2019, $r \approx 0{,}998$: ajustement excellent · résidus +2, +1, +2 · 373 k€ estimés pour 2026, seuil de 400 k€ franchi en 2028 (rang 9) · pour 2050, extrapolation à 26 ans : prévision de 829 k€ à ne pas prendre au sérieux.

Rappel de cours. Une étude complète enchaîne : point moyen, ajustement, interprétation des coefficients, résidus, prévision, seuil, puis CRITIQUE du domaine de validité.

Erreur fréquente. Oublier de reconvertir le rang en année. Le rang 9 n'est pas l'année 2009 ni 2019 : c'est $2019 + 9 = 2028$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).