Plaquette 3 – Régression linéaire et corrélation

Statistiques à deux variables — Première générale

Corrections détaillées

Boîte à outils

Régression linéaire et coefficient de corrélation.

— **Droite de régression** de y en x (méthode des moindres carrés) : la droite d'équation

$$y = ax + b$$

qui rend **minimale** la somme des carrés des résidus $\sum (y_i - (ax_i + b))^2$. Les coefficients a et b sont donnés par la calculatrice, en mode statistiques à deux variables.

— **Propriété.** Cette droite passe **toujours** par le point moyen $G(\overline{x}\,;\overline{y})$:

$$\overline{y} = a\overline{x} + b$$

C'est le contrôle à faire systématiquement après avoir relevé a et b.

— **Coefficient de corrélation linéaire.** La calculatrice affiche r (parfois r et r^2), avec

$$-1 \leq r \leq 1$$

Le **signe** de r est celui de a. Plus $|r|$ est proche de 1, meilleur est l'alignement ; $|r| = 1$ correspond à un alignement parfait.

— **Repères d'interprétation** : $|r| > 0{,}95$ ajustement affine très satisfaisant ; $0{,}8 < |r| < 0{,}95$ acceptable, à commenter ; $|r| < 0{,}8$ ajustement affine peu pertinent.

— ⚠ **Corrélation n'est pas causalité.** Un r proche de 1 dit que les deux variables varient ensemble, pas que l'une cause l'autre : il peut exister une **variable cachée** qui agit sur les deux, ou une pure coïncidence.

— **Arrondis.** On garde les coefficients avec **au moins trois chiffres significatifs** pour les calculs, et on n'arrondit qu'à la fin : un arrondi trop tôt sur a se multiplie par x et peut fausser l'estimation.

Correction de l'exercice 1 – Obtenir la droite de régression

a) La droite de régression a pour équation

$$y = ax + b \quad \text{soit} \quad y = 3{,}83x + 7{,}60$$

b) On remplace x par $\overline{x} = 3{,}5$ dans l'équation, en utilisant la valeur non arrondie de a :

$$y = 3{,}829 \times 3{,}5 + 7{,}6 \approx 13{,}40 + 7{,}6 = 21{,}00$$

On retrouve $\overline{y} = 21$: la droite passe bien par G ✓ C'est une **propriété générale** de la droite des moindres carrés, et donc un contrôle qui doit toujours fonctionner.

c) Les deux droites sont très proches :

$$\text{régression: } y = 3{,}83x + 7{,}60 \qquad \text{Mayer: } y = 3{,}78x + 7{,}77$$

Pour $x = 8$, elles donnent respectivement 38,2 et 38,0 milliers d'euros : l'écart est de 200 €, négligeable. C'est rassurant : sur un nuage bien aligné, la méthode de Mayer (calculable à la main) approche très bien la droite optimale.

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'avantage de la régression. Elle est **unique** et **optimale** : aucune autre droite ne donne une somme des carrés des résidus plus petite. La droite de Mayer, elle, dépend de la façon dont on coupe le nuage en deux.

Réponse. a) $y = 3{,}83x + 7{,}60$ — b) pour $x = 3{,}5$, on trouve $21 = \overline{y}$ ✓ — c) les deux droites sont très proches (écart de 200 € pour $x = 8$).

Correction de l'exercice 2 – Une régression décroissante

a) L'équation est

$$y = -1{,}25x + 22{,}71$$

b) Le coefficient $r \approx -0{,}999$ appelle deux commentaires distincts.

Son **signe** est négatif, comme celui de a : la corrélation est négative, donc la consommation **diminue** quand la température augmente.

Sa **valeur absolue**, 0,999, est extrêmement proche de 1 : le nuage est presque parfaitement aligné, et l'ajustement affine est **très satisfaisant** (bien au-delà du seuil de 0,95).

c) Pour $x = 18$:

$$y = -1{,}25 \times 18 + 22{,}71 = -22{,}5 + 22{,}71 = 0{,}21$$

L'estimation est de 0,21 m³ par jour. Elle est **à peine exploitable** : d'une part 18 °C est en dehors de l'intervalle des relevés (de 2 à 14 °C), d'autre part le modèle donnerait une consommation **négative** dès 19 °C, ce qui est impossible. Le modèle affine cesse d'être valable dès que le chauffage est éteint.

La leçon. Un excellent r garantit la qualité de l'ajustement **sur l'intervalle des données observées**, et rien du tout au-delà.

Réponse. a) $y = -1{,}25x + 22{,}71$ — b) signe négatif (corrélation négative), $|r|$ très proche de 1 : excellent ajustement — c) 0,21 m³/j, mais hors intervalle et bientôt négatif : peu exploitable.

Correction de l'exercice 3 – Comparer deux coefficients de corrélation

a) On compare $|r|$ aux repères du cours.

Série A : $|r| = 0{,}957 > 0{,}95$, l'ajustement affine est **pertinent** — le nuage est nettement allongé, malgré deux valeurs qui s'écartent de la tendance.

Série B : $|r| = 0{,}165$, très loin de 1. L'ajustement affine n'est **pas pertinent** : le nuage n'a pas de direction privilégiée.

b) C'est la **série A**, et de très loin : 0,957 contre 0,165.

c) Cette droite existe — la calculatrice la calcule toujours — mais elle **n'a aucune valeur prédictive**. Son coefficient directeur 0,3 est minuscule devant les écarts observés (les ordonnées vont de 2 à 9), et l'estimation qu'elle donnerait pour $x = 6$, soit $0{,}3 \times 6 + 4{,}5 = 6{,}3$, ne repose sur rien.

La règle de rédaction. On calcule **toujours** r **avant** d'exploiter une droite de régression, et on justifie la pertinence de l'ajustement avant d'estimer quoi que ce soit. Une estimation sans commentaire sur r est une réponse incomplète.

Réponse. a) A : pertinent (0,957) ; B : non pertinent (0,165) — b) la série A — c) elle existe mais n'a aucune valeur prédictive.

Correction de l'exercice 4 – Masse et taille de nouveau-nés

a) L'équation est

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).

$$y = 0{,}170x - 5{,}206$$

Le coefficient $r \approx 0{,}998$ est très proche de 1 : l'alignement est excellent, l'ajustement affine est pleinement justifié sur cet intervalle de tailles.

b) On remplace x par 49 :

$$y = 0{,}170 \times 49 - 5{,}206 = 8{,}330 - 5{,}206 = 3{,}124$$

La masse estimée est donc d'environ 3,124 kg, soit 3 124 **g**. C'est une **interpolation** (la valeur 49 est entre 48 et 50), donc une estimation fiable.

c) Le coefficient $a = 0{,}170$ s'interprète en kilogrammes **par centimètre** : chaque centimètre de taille supplémentaire correspond, en moyenne, à 0,170 kg de plus, soit 170 g.

Le coefficient $b = -5{,}206$ serait la masse estimée d'un nouveau-né de taille **nulle** : −5,206 kg, une masse négative, donc absurde. Il n'a aucune signification concrète — il ne sert qu'à **positionner** la droite dans le repère.

Le principe général. L'ordonnée à l'origine ne s'interprète que si $x = 0$ correspond à une situation réelle. Ici, les tailles observées vont de 46 à 54 cm : extrapoler jusqu'à 0 sort complètement du domaine de validité.

Réponse. a) $y = 0{,}170x - 5{,}206$, avec $r \approx 0{,}998$: excellent ajustement — b) environ 3 124 g — c) a : 170 g par centimètre ; b n'a pas de sens (masse négative pour une taille nulle).

Correction de l'exercice 5 – Quand la corrélation est parfaite

a) $r = 1$ signifie que la corrélation linéaire est **parfaite** : les cinq points sont exactement **alignés**, et la droite de régression les contient tous. Il n'y a aucun résidu, donc la somme des carrés des résidus est nulle — le minimum absolu.

b) On teste les cinq points dans l'équation $y = 3x + 7$:

— $x = 1$: $3 + 7 = 10$ ✓
— $x = 2$: $6 + 7 = 13$ ✓
— $x = 3$: $9 + 7 = 16$ ✓
— $x = 5$: $15 + 7 = 22$ ✓
— $x = 6$: $18 + 7 = 25$ ✓

Les cinq points appartiennent bien à la droite.

c) Oui, c'est exactement la même : la droite joignant (1 ; 10) et (6 ; 25) avait pour équation $y = 3x + 7$. Quand les points sont alignés, **toutes** les méthodes d'ajustement donnent la même droite — celle qui passe par tous les points. Les méthodes ne diffèrent que sur les nuages imparfaits.

Le contrôle par le point moyen. $\overline{x} = \frac{17}{5} = 3{,}4$ et $\overline{y} = \frac{86}{5} = 17{,}2$, et $3 \times 3{,}4 + 7 = 17{,}2$ ✓
La droite passe par G, comme toujours.

Réponse. a) les points sont parfaitement alignés — b) les cinq points vérifient $y = 3x + 7$ — c) oui, la même : sur un nuage aligné, toutes les méthodes coïncident.

Correction de l'exercice 6 – Surface et loyer

a) L'équation est

$$y = 13{,}34x + 217{,}91$$

Le coefficient $r \approx 0{,}9998$ est quasiment égal à 1 : le nuage est presque parfaitement aligné, l'ajustement affine est excellent.

b) On remplace x par 45 :

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).

$$y = 13{,}343 \times 45 + 217{,}905 \approx 600{,}44 + 217{,}905 \approx 818{,}34$$

Le loyer estimé est donc d'environ 818 **€** par mois. Il s'agit d'une interpolation (45 est entre 40 et 50), donc d'une estimation fiable.

c) Le coefficient $a \approx 13{,}34$ s'exprime en euros **par mètre carré** : chaque mètre carré supplémentaire coûte environ 13,34 € de loyer mensuel.

Le coefficient $b \approx 217{,}91$ est le loyer estimé pour une surface **nulle**. Il ne correspond à aucun appartement réel, mais il s'interprète tout de même comme une **partie fixe** du loyer : charges, accès, situation — ce qui se paie indépendamment de la surface. C'est pourquoi le loyer n'est pas proportionnel à la surface.

Le test de proportionnalité. Si le loyer était proportionnel à la surface, on aurait $b = 0$. Ici $\frac{480}{20} = 24$ € le m² pour le petit appartement, contre $\frac{1150}{70} \approx 16{,}43$ € pour le grand : le prix au m² baisse avec la taille, exactement parce qu'il existe une partie fixe.

Réponse. a) $y = 13{,}34x + 217{,}91$, $r \approx 0{,}9998$: excellent — b) environ 818 € — c) 13,34 € par m², plus une partie fixe d'environ 218 €.

Correction de l'exercice 7 – Associer trois nuages à trois corrélations

a) Deux critères suffisent : le **signe** de r donne le sens de la tendance, sa **valeur absolue** la qualité de l'alignement.

— Nuage n° 1 : les points montent régulièrement et sont presque alignés, donc r est positif et proche de 1 : c'est $r \approx 0{,}98$.
— Nuage n° 2 : les points descendent régulièrement et sont presque alignés, donc r est négatif et proche de -1 : c'est $r \approx -0{,}99$.
— Nuage n° 3 : les points sont dispersés, sans direction, donc $|r|$ est petit : c'est $r \approx 0{,}15$.

b) Pour le **nuage n° 3**. Avec $|r| = 0{,}15$, très loin du seuil de 0,8, la droite de régression n'aurait aucune valeur prédictive : il faut écarter l'ajustement affine, et éventuellement chercher un autre modèle.

c) Un alignement **parfait** donne $|r| = 1$, et la droite étant **descendante**, le coefficient directeur est négatif, donc r aussi :

$$r = -1$$

Le résumé en une phrase. Le signe de r est celui de la pente ; sa valeur absolue mesure la qualité de l'alignement, de 0 (aucun lien linéaire) à 1 (alignement exact).

Réponse. a) n° 1 : 0,98 ; n° 2 : −0,99 ; n° 3 : 0,15 — b) le nuage n° 3 — c) $r = -1$.

Correction de l'exercice 8 – La droite passe toujours par G

a) La droite de régression passe **toujours** par le point moyen, donc les coordonnées de G vérifient son équation :

$$\overline{y} = a\overline{x} + b$$

On remplace par les valeurs connues :

$$45 = 2{,}8 \times 12 + b \Leftrightarrow 45 = 33{,}6 + b \Leftrightarrow b = 11{,}4$$

b) L'équation de la droite de régression est donc

$$y = 2{,}8x + 11{,}4$$

Contrôle : pour $x = 12$, $y = 33{,}6 + 11{,}4 = 45 = \overline{y}$ ✓

c) Pour $x = 20$:

$$y = 2{,}8 \times 20 + 11{,}4 = 56 + 11{,}4 = 67{,}4$$

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).

Ce que l'exercice montre. La propriété « la droite passe par G » n'est pas seulement un contrôle : elle permet de **retrouver** b à partir de a et du point moyen, sans les données individuelles. C'est une question classique quand l'énoncé ne fournit que des résumés statistiques.

La réserve à formuler. L'estimation en $x = 20$ suppose que 20 reste dans ou près de l'intervalle des données observées, ce que l'énoncé ne précise pas : il faudrait le vérifier avant de conclure.

Réponse. a) $b = 11{,}4$ — b) $y = 2{,}8x + 11{,}4$ — c) environ 67,4.

Correction de l'exercice 9 – Budget publicitaire et ventes

a) L'équation est

$$y = 9{,}00x + 78{,}35$$

Avec $r \approx 0{,}999$, l'ajustement affine est excellent sur l'intervalle observé.

b) Le coefficient $a \approx 9$ s'interprète en milliers d'euros de ventes **par** millier d'euros de publicité : chaque euro investi en publicité rapporte environ 9 **€ de ventes supplémentaires**.

D'après ce modèle, augmenter le budget publicitaire est donc rentable — à condition de comparer ce gain de ventes à la **marge** réalisée dessus, et non au chiffre d'affaires : si la marge est de 10 %, 9 € de ventes ne rapportent que 0,90 € de bénéfice pour 1 € dépensé, et l'opération serait perdante.

c) Pour $x = 30$:

$$y = 8{,}999 \times 30 + 78{,}350 = 269{,}97 + 78{,}35 = 348{,}32$$

Les ventes estimées sont donc d'environ 348 **milliers d'euros**. C'est une extrapolation modérée (le budget maximal observé est 25), donc plausible mais à annoncer comme telle.

Ce que le modèle ne dit pas. Un modèle affine suppose un rendement **constant** de la publicité. En réalité, les rendements finissent toujours par décroître : à 100 milliers d'euros de budget, le modèle prédirait 978 milliers d'euros de ventes, ce qui est très douteux.

Réponse. a) $y = 9{,}00x + 78{,}35$ — b) environ 9 € de ventes par euro de publicité ; rentable seulement si la marge le justifie — c) environ 348 milliers d'euros.

Correction de l'exercice 10 – Corrélation n'est pas causalité

a) **Non.** La corrélation est positive et forte, mais ce ne sont pas les pompiers qui causent les dégâts : c'est l'**ampleur de l'incendie** qui cause à la fois la mobilisation de nombreux pompiers et les dégâts importants. L'ampleur du sinistre est une **variable cachée** qui agit sur les deux.

b) **Non.** Un enfant plus grand lit mieux, mais grandir ne fait pas lire : c'est l'**âge** qui agit sur la taille et sur le niveau de lecture. Encore une variable cachée.

c) **Plausiblement oui**, partiellement. Ici un mécanisme causal existe : l'entraînement améliore effectivement la performance. Mais le lien n'est pas exclusif — le talent, le matériel, l'encadrement, la récupération jouent aussi, et l'entraînement excessif peut même dégrader la performance.

d) **Non.** Les deux ventes augmentent ensemble, mais aucune ne cause l'autre : c'est la **température** (ou la saison) qui les fait croître simultanément. C'est l'exemple classique de corrélation fortuite par variable cachée.

La règle à retenir. Une corrélation forte n'établit **jamais** la causalité à elle seule. Ce

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).

que mesure le coefficient de corrélation, c'est uniquement la proximité du nuage avec une droite :

$$|r| \text{ proche de } 1 \Leftrightarrow \text{les points sont presque alignés}$$

et cet énoncé ne contient aucune notion de cause. Pour affirmer un lien de cause à effet, il faut un **mécanisme** identifié, et de préférence une expérience contrôlée. C'est la question de recul la plus fréquemment posée sur ce chapitre.

Réponse. a) non, variable cachée : l'ampleur de l'incendie — b) non, variable cachée : l'âge — c) oui en partie, un mécanisme existe — d) non, variable cachée : la température.

Correction de l'exercice 11 – L'effet d'un arrondi trop brutal

a) Avec les coefficients complets :

$$y = 3{,}828\,6 \times 10 + 7{,}6 = 38{,}286 + 7{,}6 = 45{,}886 \approx 45{,}9$$

L'estimation est d'environ 45,9 milliers d'euros.

b) Avec a arrondi à 4 :

$$y = 4 \times 10 + 7{,}6 = 40 + 7{,}6 = 47{,}6$$

c) L'écart entre les deux estimations est de

$$47{,}6 - 45{,}9 = 1{,}7 \text{ millier d'euros}$$

soit 1 700 €, pour une seule décimale négligée. Et cet écart **grandit avec** x : pour $x = 30$, on obtiendrait 122,5 contre 127,6, soit 5 100 € d'écart.

La règle est donc claire : on conserve a avec **au moins trois chiffres significatifs** pendant les calculs, et on n'arrondit que le **résultat final**. Arrondir a avant de multiplier par x multiplie l'erreur par x.

Le même principe qu'en pourcentages. Comme pour les coefficients multiplicateurs, l'arrondi intermédiaire est l'erreur la plus coûteuse — et la plus invisible, puisque le calcul « a l'air » juste.

Réponse. a) environ 45,9 — b) 47,6 — c) un écart de 1 700 €, qui croît avec x : on n'arrondit qu'à la fin.

Correction de l'exercice 12 – Régression sur les notes de huit élèves

a) L'équation est

$$y = 1{,}48x + 1{,}36$$

Le coefficient $r \approx 0{,}957$ dépasse le seuil de 0,95 : l'ajustement affine est **satisfaisant**, même si le nuage est visiblement moins bien aligné que dans les exercices précédents (deux élèves s'écartent nettement de la tendance).

b) On remplace x par 4,5, avec les valeurs non arrondies :

$$y = 1{,}476 \times 4{,}5 + 1{,}357 = 6{,}642 + 1{,}357 = 7{,}999 \approx 8$$

On retrouve $\overline{y} = 8$ aux arrondis près ✓

c) Pour $x = 10$:

$$y = 1{,}476 \times 10 + 1{,}357 = 14{,}76 + 1{,}357 \approx 16{,}12$$

La note estimée est d'environ 16 sur 20. Trois réserves s'imposent :

— c'est une **extrapolation** : aucun élève n'a révisé plus de 8 heures ;

— le modèle donnerait une note supérieure à 20 dès 13 heures de révision ($1{,}476 \times 13 + 1{,}357 \approx 20{,}55$), ce qui est impossible : le modèle affine n'est pas valable

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).

très longtemps ;

— il s'agit d'une tendance **moyenne** sur huit élèves, pas d'une prédiction individuelle — deux élèves du tableau s'écartent de 3 à 4 points de la droite.

La rédaction attendue. Une estimation s'accompagne toujours d'un commentaire sur sa validité : intervalle des données, plausibilité du résultat, dispersion du nuage.

Réponse. a) $y = 1{,}48x + 1{,}36$, $r \approx 0{,}957$: ajustement satisfaisant — b) on retrouve 8 ✓ — c) environ 16 / 20, mais extrapolation, note plafonnée à 20, et tendance moyenne seulement.

Correction de l'exercice 13 – Ventes de vélos : régression et Mayer

a) L'équation est

$$y = 8{,}571x + 25$$

Pour le huitième trimestre, $x = 8$:

$$y = 8{,}571 \times 8 + 25 = 68{,}57 + 25 = 93{,}57 \approx 94 \text{ vélos}$$

b) Avec la droite de Mayer, on avait obtenu $8{,}45 \times 8 + 25{,}43 = 93{,}03$, soit environ 93 vélos. L'écart entre les deux estimations est donc de moins d'un vélo.

c) Il faut retenir celle de la **droite de régression** : elle minimise la somme des carrés des résidus, c'est donc le meilleur ajustement affine possible au sens des moindres carrés. La droite de Mayer n'est qu'une approximation, obtenue en résumant le nuage par deux points.

Cela dit, l'écart est ici **négligeable** (93 contre 94 vélos) : avec $r \approx 0{,}998$, le nuage est si bien aligné que toutes les méthodes convergent. La méthode de Mayer garde donc son intérêt quand on n'a pas de calculatrice.

Le contrôle par G. $\overline{x} = 3{,}5$ et $\overline{y} = 55$. Régression : $8{,}571 \times 3{,}5 + 25 \approx 55{,}0$ ✓ Mayer : $8{,}45 \times 3{,}5 + 25{,}43 \approx 55{,}0$ ✓ Les deux droites passent par le point moyen, ce qui explique qu'elles se croisent en son voisinage.

Réponse. a) $y = 8{,}571x + 25$ et environ 94 vélos — b) Mayer donnait 93 vélos — c) la régression, qui est optimale ; mais l'écart est négligeable ici.

Correction de l'exercice 14 – Lire les sorties d'une calculatrice

a) La droite de régression a pour équation

$$y = ax + b \quad \text{soit} \quad y = -0{,}845x + 53{,}17$$

b) Le coefficient de corrélation vaut $r \approx -0{,}872$, donc $|r| \approx 0{,}872$. D'après les repères du cours, cette valeur est comprise entre 0,8 et 0,95 : l'ajustement est **acceptable mais à commenter**. Le nuage est nettement orienté à la baisse, mais les points s'écartent sensiblement de la droite — le carré $r^2 = 0{,}760$ indique que 76 % de la variabilité de y est « expliquée » par la relation affine, et 24 % ne l'est pas.

On pourra donc utiliser la droite pour dégager une tendance, en précisant que les estimations individuelles seront imprécises.

c) Oui, ils sont cohérents : a et r sont tous deux **négatifs**, et le cours affirme que le signe de r est celui de a. S'ils étaient de signes différents, ce serait nécessairement une **erreur** — de saisie des données ou de lecture de l'écran — et il faudrait tout recommencer.

Le rôle de r^2. La calculatrice l'affiche souvent à côté de r : c'est la part de variabilité expliquée par l'ajustement. Il est toujours positif, donc lui seul ne renseigne pas sur le **sens** de la corrélation : c'est r, ou le signe de a, qu'il faut regarder pour cela.

Réponse. a) $y = -0{,}845x + 53{,}17$ — b) $|r| \approx 0{,}872$, entre 0,8 et 0,95 : acceptable, à

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).

commenter — c) oui, les deux sont négatifs ; des signes opposés signaleraient une erreur.

Correction de l’exercice 15 – Synthèse : consommation d’un véhicule

a) Les deux moyennes, avec $n = 6$:

$$\overline{x} = \frac{50 + 70 + 90 + 100 + 110 + 130}{6} = \frac{550}{6} \approx 91{,}67$$

$$\overline{y} = \frac{5{,}2 + 5{,}0 + 5{,}4 + 5{,}9 + 6{,}5 + 7{,}8}{6} = \frac{35{,}8}{6} \approx 5{,}97$$

Le point moyen est environ $G(91{,}67\,;5{,}97)$.

b) L’équation est

$$y = 0{,}032\,65x + 2{,}973$$

Le coefficient $r \approx 0{,}890$ est compris entre 0,8 et 0,95 : l’ajustement affine est **acceptable**, mais il mérite un commentaire — le nuage n’est pas parfaitement aligné, et c’est le moins bon coefficient de toute la plaquette.

c) On remplace x par $\overline{x} = 91{,}67$:

$$y = 0{,}032\,65 \times 91{,}67 + 2{,}973 \approx 2{,}993 + 2{,}973 = 5{,}966$$

On retrouve $\overline{y} \approx 5{,}97$: la droite passe bien par le point moyen ✓ Le léger écart sur la troisième décimale ne vient que des arrondis de l’énoncé.

d) Pour $x = 120$:

$$y = 0{,}032\,65 \times 120 + 2{,}973 = 3{,}918 + 2{,}973 = 6{,}891$$

La consommation estimée est d’environ 6,9 L/100 km. C’est une **interpolation** (120 est entre 110 et 130), donc raisonnable.

e) Non, pas tout à fait. Les trois premières valeurs sont 5,2, puis 5,0, puis 5,4 : la consommation **baisse** d’abord entre 50 et 70 km/h, avant de remonter. Un modèle affine croissant ne peut pas rendre compte de cette baisse initiale. Physiquement, il existe une vitesse de consommation minimale (autour de 70 km/h ici), et la consommation remonte de part et d’autre : le nuage serait bien mieux décrit par une **parabole** que par une droite.

Le bilan de la plaquette. La calculatrice donne toujours une droite et un r ; c’est au rédacteur de décider si le modèle affine est légitime, en regardant le nuage **et** la valeur de r, et en interrogeant le sens physique de la situation.

Réponse. a) $G(91{,}67\,;5{,}97)$ — b) $y = 0{,}032\,65x + 2{,}973$, $r \approx 0{,}890$: acceptable — c) la droite passe par G ✓ — d) environ 6,9 L/100 km — e) non : la consommation baisse d’abord, un modèle parabolique conviendrait mieux.

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).