Plaquette 1 – Nuage de points, ajustement affine et prévisions

Statistiques à deux variables — Terminale STMG

Corrections détaillées

Boîte à outils

Méthodes et formules du cours.

— Un **nuage de points** représente les couples $(x_i\,;y_i)$ dans un repère.

— **Point moyen** G : ses coordonnées sont les moyennes des deux séries,

$$G(\overline{x}\,;\overline{y}) \qquad \text{avec} \qquad \overline{x}=\frac{x_1+\cdots+x_n}{n}$$

— La **droite d'ajustement affine** obtenue par la méthode des moindres carrés a une équation de la forme $y=ax+b$ (coefficients donnés par la calculatrice). Elle passe **toujours** par le point moyen G.

— **Interprétation de** a : quand x augmente de 1, y varie de a en moyenne.

— **Prévision** : on remplace x par la valeur souhaitée dans l'équation.

— **Interpolation** : la valeur de x est à l'intérieur de la plage observée (prévision fiable) ;

— **extrapolation** : elle est en dehors (prévision à prendre avec prudence).

— **Méthode des moindres carrés** : parmi toutes les droites $y=ax+b$, c'est celle qui rend minimale la somme des carrés des écarts verticaux

$$S=\sum_{i=1}^{n}\left(y_i-(ax_i+b)\right)^2$$

— Un ajustement affine est **pertinent** quand les points du nuage sont presque alignés ; on en juge sur le nuage (et par la petitesse de S), pas par une formule.

— Ajustement affine ne veut pas dire **causalité** : la droite décrit une tendance observée, elle n'explique pas.

Correction de l'exercice 1 – Calculer le point moyen

On calcule séparément les deux moyennes.

— $\overline{x}=\dfrac{10+12+15+18+20}{5}=\dfrac{75}{5}=15$.

— $\overline{y}=\dfrac{122+128+145+158+172}{5}=\dfrac{725}{5}=145$.

— Le point moyen est donc $G(15\,;145)$.

— Remarque : ce point appartiendra nécessairement à la droite d'ajustement.

Réponse. $G(15\,;145)$.

Correction de l'exercice 2 – Lire une équation d'ajustement

a) On remplace x par 15 :

$$5\times 15+70=75+70=145$$

— On obtient bien $\overline{y}=145$: la droite passe par $G(15\,;145)$, comme l'affirme le cours.

b) Le coefficient directeur 5 signifie que **chaque millier d'euros investi en publicité correspond, en moyenne, à** 5 **milliers d'euros de chiffre d'affaires supplémentaires** sur la période observée.

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 70 représente le chiffre d'affaires « théorique » sans publicité — à interpréter avec prudence, car $x = 0$ est loin des valeurs observées.

Réponse. a) $5 \times 15 + 70 = 145 = \overline{y}$ · b) +1 millier de publicité → +5 milliers de CA en moyenne.

Correction de l'exercice 3 – Faire une prévision

a) $y = 5 \times 17 + 70 = 85 + 70 = 155$ milliers d'euros.

b) $y = 5 \times 30 + 70 = 150 + 70 = 220$ milliers d'euros.

c) La première : 17 est **à l'intérieur** de la plage observée (10 à 20), c'est une **interpolation**.

— La seconde est une **extrapolation** : 30 est bien au-delà des données. Rien ne garantit que la tendance se poursuive (saturation du marché, budget publicitaire moins efficace…). La prévision est donc peu fiable.

Réponse. a) 155 · b) 220 · c) la première (interpolation).

Correction de l'exercice 4 – Prévision inverse

On résout l'équation $5x + 70 = 180$.

— $5x = 180 - 70 = 110$.

— $x = \dfrac{110}{5} = 22$ milliers d'euros.

— Cette valeur est juste au-delà de la plage observée (10 à 20) : l'estimation reste plausible, mais il s'agit déjà d'une légère extrapolation.

— Contrôle : $5 \times 22 + 70 = 180$.

Réponse. 22 milliers d'euros.

Correction de l'exercice 5 – Juger un nuage de points

On regarde si les points se répartissent autour d'une droite.

— **Nuage A** : les points sont presque alignés, et y augmente d'environ 2 quand x augmente de 1 (de 3,1 à 4,9, puis à 7,2…). Un ajustement affine est pertinent.

— **Nuage B** : les points dessinent une courbe tournée vers le haut. Les écarts entre deux valeurs successives augmentent :

$$0{,}5 \qquad 1{,}1 \qquad 1{,}8 \qquad 3{,}4 \qquad 4{,}7$$

Une droite décrirait mal ce nuage. On pensera plutôt à un modèle exponentiel ou à un **changement de variable** (par exemple $z = \ln(y)$), qui redresse le nuage.

Réponse. Nuage A : oui (points presque alignés) ; nuage B : non (nuage incurvé).

Correction de l'exercice 6 – Ajustement d'une série chronologique

a) $\overline{x} = \dfrac{1+2+3+4+5+6}{6} = \dfrac{21}{6} = 3{,}5$.

— $\overline{y} = \dfrac{210+231+246+266+281+302}{6} = \dfrac{1\,536}{6} = 256$.

— Point moyen : $G(3{,}5\,;256)$. Contrôle : $18 \times 3{,}5 + 193 = 256$, la droite passe bien par G.

b) 2026 correspond au rang $x = 7$.

— $y = 18 \times 7 + 193 = 126 + 193 = 319$.

— On peut estimer environ 319 **adhérents** en 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).

— Contrôle de cohérence : la tendance est de +18 adhérents par an et le dernier relevé vaut 302, donc $302 + 18 = 320$. L'ordre de grandeur est confirmé.

Réponse. a) $G(3{,}5\,;256)$ · b) environ 319 adhérents.

Correction de l'exercice 7 – Changement d'unité

a) Passer des milliers d'euros aux euros revient à multiplier toutes les valeurs de y par 1 000 : chaque coefficient est donc multiplié par 1 000.

$$y = 2\,400x + 15\,000$$

b) Oui : l'ordonnée $\overline{y}$ du point moyen est elle aussi multipliée par 1 000, et la nouvelle droite passe par ce nouveau point moyen.

— Changer d'unité étire le nuage verticalement sans modifier l'**alignement** des points : la qualité de l'ajustement est la même. Seuls les coefficients de la droite changent.

Réponse. a) $y = 2\,400x + 15\,000$ · b) oui ; la qualité de l'ajustement est inchangée.

Correction de l'exercice 8 – Critiquer un modèle

a) $y = 32 \times 25 - 180 = 620$ glaces : 25 °C est dans la plage observée, c'est une interpolation fiable.

b) $y = 32 \times 3 - 180 = 96 - 180 = -84$ glaces. Un nombre de glaces **négatif** est absurde : le modèle affine n'est valable que sur la plage de températures observée, vraisemblablement des températures estivales. En dehors, il perd tout sens.

c) Non. Un bon ajustement n'établit **pas** une causalité : ici le lien est plausible, mais l'alignement des points seul ne le démontre pas. D'autres facteurs (vacances scolaires, jour de la semaine) varient en même temps que la température.

Réponse. a) 620 glaces · b) -84 glaces : le modèle sort de son domaine de validité · c) non, un ajustement ne prouve pas une causalité.

Correction de l'exercice 9 – Comparer deux prévisions

a) $\overline{x} = \dfrac{1+2+3+4+5}{5} = 3$.

— $\overline{y} = \dfrac{45+52+56+63+69}{5} = \dfrac{285}{5} = 57$.

— Point moyen : $G(3\,;57)$. Contrôle : $5{,}9 \times 3 + 39{,}3 = 17{,}7 + 39{,}3 = 57$. La droite passe bien par G.

b) Rang 6 : $5{,}9 \times 6 + 39{,}3 = 35{,}4 + 39{,}3 = 74{,}7$ milliers d'unités.

— Rang 12 : $5{,}9 \times 12 + 39{,}3 = 70{,}8 + 39{,}3 = 110{,}1$ milliers d'unités.

c) Celle de rang 6 : elle prolonge d'un an seulement la période observée.

— La prévision de rang 12 suppose que la tendance linéaire se maintienne pendant 7 ans de plus, ce qui est très incertain (cycle de vie du produit, concurrence). À mentionner, mais avec une réserve explicite.

Réponse. a) $G(3\,;57)$ · b) 74,7 puis 110,1 · c) celle de rang 6.

Correction de l'exercice 10 – Ajustement après transformation

— a) Une croissance de 10 % par an signifie que y est **multiplié** par 1,10 chaque année : les écarts augmentent d'année en année. Un nuage de points prendrait une allure incurvée vers le haut, et une droite le décrirait mal : elle sous-estimerait les valeurs récentes et surestimerait les anciennes.

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).

— b) Il faut un modèle **exponentiel** : $y = y_0 \times 1{,}10^{x}$, cohérent avec une évolution en pourcentage constant.
— Indicateur de contrôle : si les quotients z d'une année sur l'autre sont à peu près constants, le modèle exponentiel convient ; si ce sont les **différences** qui sont constantes, le modèle affine convient.

Réponse. a) les écarts croissent, le nuage est incurvé · b) un modèle exponentiel $y = y_0 \times 1{,}10^{x}$.

Correction de l'exercice 11 – Une droite au jugé

a) Le coefficient directeur vaut

$$a = \frac{302 - 256}{6 - 3{,}5} = \frac{46}{2{,}5} = 18{,}4$$

et $b = 256 - 18{,}4 \times 3{,}5 = 256 - 64{,}4 = 191{,}6$: la droite au jugé est $y = 18{,}4x + 191{,}6$.

b) Au rang 8 : $18{,}4 \times 8 + 191{,}6 = 338{,}8$ avec la droite au jugé, et $18 \times 8 + 193 = 337$ avec les moindres carrés.

Les deux prévisions diffèrent de moins de 2 adhérents : une droite tracée au jugé en passant par G donne souvent un résultat proche de la droite des moindres carrés.

Réponse. a) $y = 18{,}4x + 191{,}6$ — b) 338,8 contre 337.

Correction de l'exercice 12 – Calculer des écarts

On calcule les valeurs données par la droite, puis les écarts :

Rang	1	2	3	4	5	6
Droite	211	229	247	265	283	301
Écart	−1	2	−1	1	−2	1

La somme des écarts vaut

$$-1 + 2 - 1 + 1 - 2 + 1 = 0$$

Les points sont répartis de part et d'autre de la droite, et les écarts se compensent : c'est une conséquence du fait que la droite passe par le point moyen. Les écarts restent petits (au plus 2 adhérents), signe d'un bon ajustement.

Réponse. Écarts -1 ; 2 ; -1 ; 1 ; -2 ; 1, de somme nulle.

Correction de l'exercice 13 – Le critère des moindres carrés

a) Pour D_1, les écarts sont 0 ; 0 ; 1 ; 0, donc $S_1 = 1$. Pour D_2, les écarts sont 0 ; $-0{,}5$; 0 ; $-1{,}5$, donc $S_2 = 0{,}25 + 2{,}25 = 2{,}5$.

b) Pour la droite des moindres carrés, les écarts sont $-0{,}1$; $-0{,}2$; $0{,}7$; $-0{,}4$:

$$S = 0{,}01 + 0{,}04 + 0{,}49 + 0{,}16 = 0{,}7$$

C'est la plus petite des trois valeurs : la droite des moindres carrés est, par définition, celle qui rend S minimale. D_1 passe exactement par trois points, mais l'écart de 1 sur le troisième pèse lourd une fois élevé au carré.

Réponse. a) $S_1 = 1$; $S_2 = 2{,}5$ — b) $S = 0{,}7$, la plus petite.

Correction de l'exercice 14 – Automatiser le calcul de S

a) La variable *total* est un **accumulateur** : initialisée à 0, elle reçoit à chaque tour de boucle le carré d'un écart. À la fin, elle contient la somme S.

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).

b) Les écarts sont 0, 0, 1 et 0 :

$$S(2\,;1) = 0 + 0 + 1 + 0 = 1$$

c) On calcule S pour chaque droite proposée (par exemple pour des valeurs de a et b obtenues par balayage) et l'on garde celle qui donne la plus petite somme : c'est le principe de la méthode des moindres carrés.

Réponse. a) un accumulateur — b) 1 — c) garder la droite de plus petite somme S.

Correction de l'exercice 15 – Synthèse : étude complète

a) $\overline{x} = \dfrac{21}{6} = 3{,}5$.

— $\overline{y} = \dfrac{84 + 79 + 77 + 72 + 68 + 65}{6} = \dfrac{445}{6} \approx 74{,}17$ €.

— Point moyen : $G(3{,}5\,;\approx 74{,}17)$.

b) Le coefficient $-3{,}8$ est **négatif** : le prix baisse en moyenne de 3,80 € par semestre sur la période observée.

c) Rang 8 : $y = -3{,}8 \times 8 + 87{,}47 = -30{,}4 + 87{,}47 = 57{,}07$ €.

— On résout $-3{,}8x + 87{,}47 < 50$: $-3{,}8x < -37{,}47$, donc $x > \dfrac{37{,}47}{3{,}8} \approx 9{,}86$.

— Le modèle prévoit un prix inférieur à 50 € à partir du rang 10.

— Réserve : au rang 10, on extrapole de quatre semestres ; par ailleurs un prix ne peut pas décroître indéfiniment (coût de production plancher).

Réponse. a) $G(3{,}5\,;\approx 74{,}17)$ · b) $-3{,}80$ € par semestre · c) 57,07 € au rang 8 ; sous 50 € à partir du rang 10.

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).