Plaquette 2 – Nuage de points et point moyen

Statistiques à deux variables — Terminale STMG

Corrections détaillées

Boîte à outils

Le nuage de points. On représente les couples $(x_i\,;y_i)$ dans un repère : x est la variable **explicative**, y la variable **expliquée**.

Le point moyen.

$$G(\overline{x}\,;\overline{y}) \qquad \overline{x}=\frac{x_1+\cdots+x_n}{n} \qquad \overline{y}=\frac{y_1+\cdots+y_n}{n}$$

La droite d'ajustement par les moindres carrés passe **toujours** par G.

Juger à l'œil.

Allure du nuage	Ajustement affine
points alignés ou presque	pertinent
forme en courbe	non pertinent : changer de variable
nuage sans forme	non pertinent

Indice utile. Pour x régulièrement espacés, des écarts $y_{i+1}-y_i$ à peu près constants signalent un nuage presque aligné.

Interpréter $y=ax+b$. Quand x augmente de 1, y varie en moyenne de a ; b est la valeur donnée par la droite pour $x=0$.

⚠ Un bon ajustement décrit une tendance conjointe ; il ne prouve **aucune** causalité.

Correction de l'exercice 1 – Calculer un point moyen

On calcule séparément la moyenne de chaque série.

$$\overline{x}=\frac{1+2+3+4+5}{5}=\frac{15}{5}=3 \qquad \overline{y}=\frac{42+48+57+63+70}{5}=\frac{280}{5}=56$$

Le point moyen est $G(3\,;56)$.

On remarque que $\overline{y}=56$ n'est aucune des valeurs du tableau : le point moyen n'est en général pas un point du nuage, c'est son « centre de gravité ».

Réponse. $G(3\,;56)$.

Correction de l'exercice 2 – Un nuage à décrire

Les abscisses augmentent de 1 à chaque fois ; on calcule les écarts entre ordonnées successives :

$$48-42=6 \qquad 57-48=9 \qquad 63-57=6 \qquad 70-63=7$$

Ces écarts sont proches les uns des autres (entre 6 et 9, autour de 7) : y augmente d'une quantité à peu près constante. Les points sont donc presque alignés, et un ajustement affine est pertinent.

La pente devrait être voisine de 7 ; la méthode des moindres carrés donne d'ailleurs $y=7{,}1x+34{,}7$.

Réponse. Oui : les écarts successifs (6, 9, 6, 7) sont proches, les points sont presque

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).

alignés.

Correction de l’exercice 3 – Trois nuages à comparer

On calcule les écarts successifs de chaque série.

— A : 4 ; 5 ; 4 ; 5. Écarts presque constants et positifs : nuage presque aligné, liaison **croissante**.
— B : −6 ; 8 ; −6 ; 5. Écarts de signes alternés : le nuage n’a pas de forme, aucun ajustement affine n’a de sens.
— C : −7 ; −6 ; −7 ; −6. Écarts presque constants et négatifs : nuage presque aligné, liaison **décroissante**.

L’ajustement affine est pertinent pour A et C, pas pour B.

$$\text{A : pente} \approx 4{,}5 \qquad \text{C : pente} \approx -6{,}5$$

Réponse. A (croissante) et C (décroissante) ; pas B.

Correction de l’exercice 4 – Point moyen d’une série chronologique

Les rangs sont $x_i = 0 ; 1 ; 2 ; 3 ; 4 ; 5$.

$$\overline{x} = \frac{0+1+2+3+4+5}{6} = \frac{15}{6} = 2{,}5 \qquad \overline{y} = \frac{45+52+56+63+71+78}{6} = \frac{365}{6} \approx 60{,}83$$

Le point moyen est $G(2{,}5\,;\approx 60{,}83)$.

L’abscisse 2,5 correspond au milieu de la période, entre 2020 et 2021 : pour des années consécutives, $\overline{x}$ est toujours le milieu de la plage.

Réponse. $G(2{,}5\,;\approx 60{,}83)$.

Correction de l’exercice 5 – Lire une pente

a) Le coefficient directeur 7,1 est positif : d’un mois au suivant, le nombre de commandes augmente en moyenne de 7,1.

b) 34,7 est la valeur donnée par la droite pour $x = 0$, c’est-à-dire le mois précédant le premier relevé. C’est une valeur théorique, à interpréter avec prudence.

c) On remplace x par 6 :

$$y = 7{,}1 \times 6 + 34{,}7 = 42{,}6 + 34{,}7 = 77{,}3$$

La droite prévoit environ 77 commandes au sixième mois.

Réponse. a) +7,1 commandes par mois en moyenne — b) valeur prédite pour $x = 0$ — c) 77,3.

Correction de l’exercice 6 – Une liaison décroissante

a) On calcule les deux moyennes :

$$\overline{x} = \frac{150}{5} = 30 \qquad \overline{y} = \frac{180+165+152+138+125}{5} = \frac{760}{5} = 152$$

d’où $G(30\,;152)$.

b) Entre le premier et le dernier point : $\dfrac{125-180}{50-10} = \dfrac{-55}{40} \approx -1{,}38$.

La pente est négative : quand le prix augmente de 1 €, les ventes baissent d’environ 1,4 unité. Les écarts successifs (−15, −13, −14, −13) sont réguliers : la liaison décroissante est très nette. La méthode des moindres carrés donne une pente de −1,37, très proche.

Réponse. a) $G(30\,;152)$ — b) pente $\approx -1{,}38$: liaison décroissante régulière.

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 7 – Reconnaître une forme

On compare les écarts et les quotients successifs.

a) Les écarts valent tous 10 : les points sont exactement alignés sur la droite $y = 10x$. L'ajustement affine est parfait.

b) Les écarts (2, 4, 8) doublent, mais les quotients sont constants :

$$\frac{4}{2} = \frac{8}{4} = \frac{16}{8} = 2$$

La série est géométrique : le nuage est courbé, un modèle exponentiel $y = 2^x$ convient, pas une droite.

c) Les écarts (3, −1, 3) sont irréguliers : la tendance est croissante, mais l'ajustement affine est médiocre.

Réponse. a) oui, parfaitement — b) non, modèle exponentiel — c) médiocre, série irrégulière.

Correction de l'exercice 8 – Ajustement et causalité

a) Non. Un nuage presque aligné montre que les deux grandeurs varient **ensemble**, pas que l'une est la cause de l'autre.

b) Un troisième facteur explique les deux : la **chaleur**. Les jours chauds, il y a plus de baigneurs, donc plus d'interventions, et plus de glaces vendues.

Un ajustement affine sert à **décrire** une tendance et à faire des prévisions sur la plage observée ; il n'explique pas les mécanismes. Avec une droite $y = ax + b$ de pente $a > 0$, on pourrait prévoir le nombre d'interventions à partir des ventes de glaces, sans que cela ait un sens pour agir : interdire les glaces ne réduirait pas les noyades.

Réponse. Non : la chaleur explique les deux phénomènes ; un ajustement ne prouve pas une causalité.

Correction de l'exercice 9 – Prévoir avec le point moyen

a) La droite s'écrit $y = 2{,}5x + b$ et passe par G :

$$45 = 2{,}5 \times 8 + b \Leftrightarrow b = 45 - 20 = 25$$

L'équation est $y = 2{,}5x + 25$.

b) Pour $x = 12$: $y = 2{,}5 \times 12 + 25 = 30 + 25 = 55$.

Le point moyen suffit, avec la pente, pour retrouver toute la droite : c'est pourquoi on le calcule en premier.

Réponse. a) $y = 2{,}5x + 25$ — b) $y = 55$.

Correction de l'exercice 10 – Interpolation ou extrapolation ?

On compare la valeur de x à la plage observée $[2\,;12]$.

a) $x = 7$ est dans la plage : c'est une **interpolation**, la prévision est fiable si le nuage est bien aligné.

b) $x = 13$ est juste au-delà : c'est une **extrapolation** proche, acceptable avec prudence.

c) $x = 40$ est très loin des données :

$$40 - 12 = 28$$

soit plus du double de l'étendue observée (10). Cette extrapolation lointaine est peu fiable : rien ne dit que la tendance se maintient aussi loin.

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. a) interpolation, fiable — b) extrapolation proche, acceptable — c) extrapolation lointaine, peu fiable.

Correction de l'exercice 11 – Repérer une erreur

Le point moyen n'est pas le point central du tableau : ses coordonnées sont les **moyennes** des deux séries.

L'abscisse est juste par hasard ($\overline{x} = 3$, car les x_i sont régulièrement espacés). Pour l'ordonnée :

$$\overline{y} = \frac{42 + 48 + 57 + 63 + 70}{5} = \frac{280}{5} = 56 \neq 57$$

Le point moyen est $G(3\,;56)$. La valeur centrale 57 serait la **médiane** des y_i, pas leur moyenne.

Réponse. $G(3\,;56)$: on calcule les moyennes, pas la valeur centrale.

Correction de l'exercice 12 – Deux séries chronologiques

a) Série A : les écarts valent tous 3, c'est une suite **arithmétique** de raison 3.

Série B : les quotients sont constants,

$$\frac{18}{12} = \frac{27}{18} = \frac{40{,}5}{27} = \frac{60{,}75}{40{,}5} = 1{,}5$$

c'est une suite **géométrique** de raison 1,5.

b) Seule A se prête à un ajustement affine (ses points sont exactement alignés). Pour B, le nuage est courbé : on utiliserait un modèle exponentiel, ou le changement de variable $z = \log(y)$, qui donne des points alignés.

Réponse. a) A arithmétique, B géométrique de raison 1,5 — b) seule A.

Correction de l'exercice 13 – Une valeur aberrante

a) $\overline{x} = 3$ et

$$\overline{y} = \frac{20 + 23 + 26 + 29 + 80}{5} = \frac{178}{5} = 35{,}6$$

d'où $G(3\,;35{,}6)$.

b) Les quatre premiers points sont exactement alignés (écarts de 3), mais le dernier saute de 29 à 80 : c'est une **valeur aberrante** (erreur de saisie, événement exceptionnel). Elle tire le point moyen vers le haut : sans elle, $\overline{y}$ vaudrait 24,5. Avant d'ajuster, il faut vérifier cette valeur et, si elle est exceptionnelle, l'écarter en le signalant.

Réponse. a) $G(3\,;35{,}6)$ — b) la valeur 80 est aberrante et fausse l'ajustement.

Correction de l'exercice 14 – Choisir la variable explicative

a) On place en abscisse la variable **explicative**, celle que l'on choisit ou que l'on connaît d'abord : la **surface** (x). Le chiffre d'affaires est la variable expliquée (y).

b) Si la droite s'écrit $y = ax + b$, la pente a est le chiffre d'affaires supplémentaire, en moyenne, pour 1 m² de surface en plus. Par exemple, avec $a = 2{,}5$:

$$2{,}5 \text{ milliers d'euros} = 2500 \text{ € par m}^2 \text{ et par an}$$

C'est un indicateur utile pour décider d'un agrandissement.

Réponse. a) la surface — b) le chiffre d'affaires gagné par m² supplémentaire.

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 15 – Estimer une pente à la main

Avec le premier et le dernier point :

$$\frac{336-250}{5-0}=\frac{86}{5}=17{,}2$$

Les écarts successifs sont 18 ; 17 ; 14 ; 19 ; 18, de moyenne $\frac{86}{5}=17{,}2$: les deux méthodes donnent le même résultat, car la somme des écarts est l’écart total.

La méthode des moindres carrés donne $\approx 16{,}97$, très proche. L’estimation à la main suffit pour un ordre de grandeur ; la calculatrice tient compte de tous les points de façon équilibrée.

Réponse. 17,2 par les deux méthodes, contre $\approx 16{,}97$ par les moindres carrés.

Correction de l’exercice 16 – Synthèse : un nuage complet

a) $\overline{x}=\frac{21}{6}=3{,}5$ et $\overline{y}=\frac{124}{6}\approx 20{,}67$: $G(3{,}5\,;\approx 20{,}67)$.

b) La pente estimée vaut $\frac{30-12}{6-1}=\frac{18}{5}=3{,}6$.

c) La droite passe par G :

$$b=\overline{y}-3{,}6\times 3{,}5\approx 20{,}67-12{,}6=8{,}07$$

soit $y=3{,}6x+8{,}07$. Pour $x=8$: $y\approx 3{,}6\times 8+8{,}07=36{,}87$. La méthode des moindres carrés donne ici exactement la même pente, 3,6.

Réponse. a) $G(3{,}5\,;\approx 20{,}67)$ — b) 3,6 — c) $y=3{,}6x+8{,}07$ et $y(8)\approx 36{,}87$.

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).