

Plaquette 3 - Méthode des moindres carrés et calculatrice

Séries statistiques à deux variables — Première STMG

Corrections détaillées

Boîte à outils

Écart (ou résidu) d'un point $M_i(x_i; y_i)$ à la droite $y = ax + b$: la différence entre la valeur observée et la valeur estimée,

$$e_i = y_i - (ax_i + b)$$

positif si le point est au-dessus de la droite, négatif s'il est en dessous.

Somme des carrés des écarts.

$$S = e_1^2 + e_2^2 + \dots + e_n^2$$

Plus S est petite, plus la droite est proche des points. On élève au carré pour que les écarts positifs et négatifs ne se compensent pas.

Droite des moindres carrés. C'est la droite qui rend S la plus petite possible. La calculatrice (mode statistiques, « régression linéaire » ou « LinReg ») ou le tableur (fonction DROITEREG, courbe de tendance) en donne les coefficients a et b . Elle passe par le point moyen G .

À la calculatrice. Saisir les x_i dans une liste, les y_i dans une autre, puis lancer la régression linéaire **de y en fonction de x** : l'ordre des listes compte.

Correction de l'exercice 1 - Calculer des écarts

a) On calcule $62x_i + 138$ pour chaque âge, puis l'écart $e_i = y_i - (62x_i + 138)$:

x_i	1	2	3	4	5
estimation	200	262	324	386	448
écart e_i	0	-2	6	-6	2

La machine de 3 ans coûte 6 € de plus que prévu par la droite (écart positif, point au-dessus) ; celle de 4 ans, 6 € de moins.

b) On additionne les carrés des écarts :

$$S = 0^2 + (-2)^2 + 6^2 + (-6)^2 + 2^2 = 0 + 4 + 36 + 36 + 4 = 80$$

Les écarts positifs et négatifs se compensent presque (leur somme vaut 0), c'est pourquoi on additionne leurs carrés pour mesurer la qualité de l'ajustement.

Réponse. a) écarts 0 ; -2 ; 6 ; -6 ; 2 — b) $S = 80$.

Correction de l'exercice 2 - Quelle droite est la meilleure ?

On calcule les écarts $e_i = y_i - (ax_i + b)$, puis $S = \sum e_i^2$.

a) Pour D_1 , les estimations sont 210, 270, 330, 390, 450 ; les écarts -10, -10, 0, -10, 0 :

$$S_1 = 100 + 100 + 0 + 100 + 0 = 300$$

Pour D_2 , les estimations sont 195, 260, 325, 390, 455 ; les écarts 5, 0, 5, -10, -5 :

$$S_2 = 25 + 0 + 25 + 100 + 25 = 175$$

b) D_2 est meilleure que D_1 , mais la droite $y = 62x + 138$, avec $S = 80$, est meilleure que les deux. C'est la droite des **moindres carrés** : aucune autre droite ne donne une somme plus petite.

Réponse. a) $S_1 = 300$, $S_2 = 175$ — b) $y = 62x + 138$ est la meilleure ($S = 80$).

Correction de l'exercice 3 - Utiliser la droite de la calculatrice

a) On remplace x par 4,5 :

$$y = 3,64 \times 4,5 + 15,93 = 16,38 + 15,93 = 32,31$$

Les ventes estimées sont d'environ 32 310 €.

b) Avec la droite de Mayer : $3,75 \times 4,5 + 15,5 = 32,375$, soit environ 32 375 €.

Les deux estimations diffèrent de 65 € à peine : pour un nuage presque aligné, les différentes méthodes d'ajustement donnent des droites voisines. Les écarts entre méthodes deviennent importants quand on s'éloigne des données (extrapolation lointaine) ou quand le nuage est plus dispersé.

Réponse. a) environ 32 310 € — b) Mayer : environ 32 375 €, presque identique.

Correction de l'exercice 4 - La droite des moindres carrés passe par G

On remplace l'abscisse de G dans l'équation et on compare avec son ordonnée.

a) $62 \times 3 + 138 = 186 + 138 = 324$: la droite passe par $G(3 ; 324)$.

b) $-20 \times 15 + 700 = -300 + 700 = 400$: la droite passe par $G(15 ; 400)$.

C'est une propriété générale de la droite des moindres carrés, admise : elle passe toujours par le point moyen. Elle donne un moyen de **contrôler** une saisie à la calculatrice : si la droite obtenue ne passe pas par G (aux arrondis près), une donnée a sans doute été mal recopiée.

Réponse. a) $62 \times 3 + 138 = 324$ — b) $-20 \times 15 + 700 = 400$.

Correction de l'exercice 5 - Prévoir un chiffre d'affaires

a) 2025 correspond au rang 7 :

$$y = 7,11 \times 7 + 112,6 = 49,77 + 112,6 = 162,37$$

Le chiffre d'affaires de 2025 est estimé à environ 162 370 €.

b) Avec la droite de Mayer : $7 \times 7 + 113 = 162$, soit 162 000 €. L'écart entre les deux prévisions est de 370 €, à peine 0,2 %. Les deux droites ont des pentes voisines (7,11 et 7) : elles traduisent la même tendance, une hausse d'environ 7 000 € par an. Il s'agit d'une extrapolation : elle suppose que cette tendance se maintient en 2025.

Réponse. a) environ 162 370 € — b) Mayer : 162 000 €, prévisions très proches.

Correction de l'exercice 6 - Âge et fréquence cardiaque maximale

a) $\bar{x} = \frac{200}{5} = 40$ et $\bar{y} = \frac{199 + 191 + 180 + 171 + 160}{5} = \frac{901}{5} = 180,2$. Donc $G(40 ; 180,2)$.
 $-0,98 \times 40 + 219,4 = -39,2 + 219,4 = 180,2$

G appartient bien à la droite.

b) Pour $x = 45$: $y = -0,98 \times 45 + 219,4 = -44,1 + 219,4 = 175,3$, soit environ 175 battements par minute. La règle usuelle donne $220 - 45 = 175$: les deux résultats sont presque identiques, ce qui confirme que cette règle est un ajustement affine de coefficient directeur -1 .

Réponse. a) $G(40; 180,2)$, sur la droite — b) environ 175 battements par minute, comme la règle.

Correction de l'exercice 7 - Visiteurs et commandes

a) Quand le nombre de visiteurs augmente de 1 millier, le nombre de commandes augmente en moyenne de 4,48 : environ 4,5 commandes pour 1 000 visiteurs supplémentaires, soit un taux de commande d'environ 0,45 % sur ces visiteurs en plus.

b) On remplace x par 22 :

$$y = 4,48 \times 22 + 5,4 = 98,56 + 5,4 = 103,96$$

On peut prévoir environ 104 commandes. C'est une interpolation, puisque 22 est compris entre 10 et 30 : l'estimation est fiable si les conditions de vente restent les mêmes.

Réponse. a) environ 4,5 commandes de plus par millier de visiteurs — b) environ 104 commandes.

Correction de l'exercice 8 - Deux méthodes, deux estimations

Pour 900 m², $x = 9$:

$$\text{Mayer: } 35 \times 9 + 85 = 400 \quad \text{moindres carrés: } 35,09 \times 9 + 84,4 = 400,21$$

Les deux estimations sont pratiquement égales : environ 400 000 €.

Pour 3 000 m², $x = 30$: Mayer donne $35 \times 30 + 85 = 1135$, les moindres carrés $35,09 \times 30 + 84,4 = 1137,1$. L'écart reste faible, mais ces deux estimations sont des **extrapolations** très lointaines : les données s'arrêtent à 1 200 m². Rien ne garantit qu'un magasin deux fois et demie plus grand suive la même tendance ; ces chiffres sont à prendre avec beaucoup de prudence.

Réponse. 900 m² : environ 400 000 € par les deux méthodes ; 3 000 m² : environ 1,14 million d'euros, extrapolation peu fiable.

Correction de l'exercice 9 - Chauffage : moindres carrés

a) $C = -3,94 \times 3 + 93,71 = -11,82 + 93,71 = 81,89$, soit environ 82 kWh par jour.

b) Avec Mayer : $-4 \times 3 + 94 = 82$ kWh. Les deux estimations coïncident à 0,11 kWh près : elles sont cohérentes.

c) On remplace T par 5 :

$$-3,94 \times 5 + 93,71 = -19,7 + 93,71 = 74,01$$

On retrouve 74 à 0,01 près : l'écart vient seulement de l'arrondi des coefficients au centième.

Réponse. a) environ 82 kWh — b) oui, Mayer donne aussi 82 kWh — c) $74,01 \approx 74$.

Correction de l'exercice 10 - Comparer deux droites sur quatre points

On calcule $e_i = y_i - (ax_i + b)$ pour chaque point, puis $S = \sum e_i^2$.

a) Pour D_1 , les estimations sont 3, 5, 7, 9 et les écarts 0, 0, -1, -1 :

$$S_1 = 0 + 0 + 1 + 1 = 2$$

Pour D_2 , les estimations sont 3,1, 4,7, 6,3, 7,9 et les écarts -0,1, 0,3, -0,3, 0,1 :

$$S_2 = 0,01 + 0,09 + 0,09 + 0,01 = 0,2$$

b) S_2 est dix fois plus petite : c'est D_2 , la droite des moindres carrés. Pourtant D_1 passe exactement par deux des points ; mais elle s'écarte des deux autres. Point moyen :

$G(2,5 ; 5,5)$, et $1,6 \times 2,5 + 1,5 = 5,5$.

Réponse. a) $S_1 = 2$, $S_2 = 0,2$ — b) D_2 , qui passe par $G(2,5 ; 5,5)$.

Correction de l'exercice 11 - Interpréter un écart

L'écart est la différence entre la valeur observée et la valeur estimée par la droite :

$$e = y - (3,64x + 15,93)$$

Avril : estimation $3,64 \times 5 + 15,93 = 34,13$, donc $e = 35 - 34,13 = 0,87$. Les ventes ont dépassé la tendance de 870 € : avril a été un **bon** mois, compte tenu de sa publicité.

Mars : estimation $3,64 \times 3 + 15,93 = 26,85$, donc $e = 26 - 26,85 = -0,85$. Les ventes sont inférieures de 850 € à la tendance.

Un écart positif place le point au-dessus de la droite, un écart négatif en dessous. Le gestionnaire peut chercher ce qui explique ces écarts (météo, jours fériés, promotions).

Réponse. Avril : $+0,87$ (au-dessus de la tendance) ; mars : $-0,85$ (en dessous).

Correction de l'exercice 12 - Atteindre un objectif de commandes

On résout l'inéquation $4,48x + 5,4 \geq 150$:

$$4,48x \geq 144,6 \Leftrightarrow x \geq \frac{144,6}{4,48} \approx 32,28$$

Il faut environ 32 300 visiteurs par jour, d'après le modèle.

Contrôle : pour $x = 33$, $4,48 \times 33 + 5,4 = 153,24$, au-dessus de 150 ; pour $x = 32$, on trouve 148,76, en dessous.

Prudence : les données s'arrêtent à 30 000 visiteurs. Cette estimation est une légère extrapolation : elle suppose que le taux de commande reste le même avec plus de visiteurs.

Réponse. Au moins 32 300 visiteurs environ.

Correction de l'exercice 13 - Repérer une erreur de saisie

L'erreur. La calculatrice ajuste toujours la seconde liste en fonction de la première. En inversant les listes, l'élève a obtenu la droite qui donne la **température en fonction de la consommation** :

$$T = -0,25C + 23,74$$

Remplacer x par 3 revient à chercher la température pour une consommation de 3 kWh, ce qui n'a rien à voir avec la question. Le résultat, 23 kWh par 3 °C, est d'ailleurs absurde : les données montrent environ 80 kWh à cette température.

La correction. On saisit les températures en première liste (x) et les consommations en seconde (y). On obtient $C = -3,94T + 93,71$, d'où $C \approx 81,89$ kWh par 3 °C.

Réponse. Les listes sont inversées : il faut ajuster C en fonction de T , et l'on trouve environ 82 kWh.

Correction de l'exercice 14 - Arrondir les coefficients

a) Pour $x = 6$:

$$4 \times 6 + 16 = 40 \quad 3,64 \times 6 + 15,93 = 37,77$$

La donnée observée pour $x = 6$ est 38. La droite trop arrondie surestime de 2 milliers d'euros, la droite arrondie au centième se trompe de 0,23 seulement.

b) Arrondir 3,64 à 4, c'est modifier la pente d'environ 10 % : toutes les estimations s'en trouvent faussées, d'autant plus que x est grand. On garde les coefficients au **centième** (ou à la précision demandée par l'énoncé), et on n'arrondit que le résultat final.

Réponse. a) 40 contre 37,77 (donnée : 38) — b) garder les coefficients au centième.

Correction de l'exercice 15 - Synthèse : remplacer une machine

a) Chaque année d'âge supplémentaire augmente le coût annuel d'entretien d'environ 62 €.

b) $y = 62 \times 7 + 138 = 434 + 138 = 572$: environ 572 € par an.

c) On résout $62x + 138 > 500$:

$$62x > 362 \Leftrightarrow x > \frac{362}{62} \approx 5,84$$

La machine sera remplacée à partir de 6 ans. Contrôle : $62 \times 5 + 138 = 448$ et $62 \times 6 + 138 = 510$.

d) Les âges 6 et 7 dépassent les données (de 1 à 5 ans) : ce sont des extrapolations. Les machines âgées peuvent tomber en panne plus souvent qu'une droite ne le prévoit ; il faudrait vérifier la tendance avec des données sur des machines plus anciennes.

Réponse. a) +62 € par an d'âge — b) environ 572 € — c) à partir de 6 ans — d) extrapolation au-delà de 5 ans.