

Plaquette 5 - Problèmes complets d'ajustement

Séries statistiques à deux variables — Terminale STI2D

Corrections détaillées

Boîte à outils

La démarche complète, en six étapes.

1. **Représenter** ou décrire le nuage. 2. **Tester** : différences, quotients, produits xy . 3. **Transformer** si nécessaire ($z = \log y$, $z = \frac{1}{y}$, $t = \sqrt{x} \dots$). 4. **Vérifier l'alignement** après transformation. 5. **Ajuster** puis **revenir** à la grandeur d'origine. 6. **Prévoir**, en distinguant interpolation et extrapolation.

Aide-mémoire des retours.

Ajustement obtenu	Modèle en y
$z = ax + b$ avec $z = \log y$	$y = 10^b \times (10^a)^x$
$z = ax + b$ avec $z = \frac{1}{y}$	$y = \frac{1}{ax + b}$
$z = ax + b$ avec $z = y^2$	$y = \sqrt{ax + b}$
$y = at + b$ avec $t = \sqrt{x}$	$y = a\sqrt{x} + b$

Qualifier une prévision.

Position de x	Terme	Fiabilité
dans la plage mesurée	interpolation	bonne
un peu au-delà	extrapolation courte	acceptable, à justifier
très au-delà	extrapolation lointaine	à éviter, confronter au bon sens

Les retours les plus fréquents.

$$\log y = ax + b \Leftrightarrow y = 10^b \times (10^a)^x \quad \frac{1}{y} = ax + b \Leftrightarrow y = \frac{1}{ax + b}$$

Le contrôle final : le résultat est-il **physiquement possible** ? Une durée négative, une concentration supérieure à 100 %, une consommation qui explose : autant de signaux d'un modèle poussé hors de son domaine.

△ Un ajustement décrit ce qu'on a mesuré, jamais ce qui **causera** quoi que ce soit.

Correction de l'exercice 1 - Une consommation croissante

a) Les différences valent 56, 60, 64 et 69 : elles augmentent, le nuage n'est pas aligné. Les quotients valent environ 1,07 à chaque fois ($\frac{856}{800} = 1,07$, $\frac{916}{856} \approx 1,070$, etc.) : le modèle est **exponentiel**.

b) On multiplie par 1,07 chaque mois, à partir de 800 :

$$y = 800 \times 1,07^x$$

soit une hausse de 7 % **par mois**. **Contrôle** en $x = 4$: $800 \times 1,3108 \approx 1\,049$ kWh. Sur un an, ce rythme multiplierait la consommation par $1,07^{12} \approx 2,25$, soit +125 %.

Réponse. a) exponentielle · b) $y = 800 \times 1,07^x$, soit +7 % par mois.

Correction de l'exercice 2 - Exploiter ce modèle

a) $y(8) = 800 \times 1,07^8 \approx 800 \times 1,718186 \approx 1\,375$ kWh. C'est une extrapolation courte, quatre mois après la dernière mesure : acceptable.

b) On résout $800 \times 1,07^x = 2\,000$, soit $1,07^x = 2,5$:

$$x = \frac{\log 2,5}{\log 1,07} \approx \frac{0,39794}{0,029384} \approx 13,54$$

La consommation dépasse 2 000 kWh au **quatorzième mois** ($y(13) \approx 1\,928$ et $y(14) \approx 2\,063$). Une hausse de 7 % par mois pendant plus d'un an est peu vraisemblable : il faudrait de nouvelles mesures avant de conclure.

Réponse. a) environ 1 375 kWh · b) au 14^e mois, sous réserve que le rythme se maintienne.

Correction de l'exercice 3 - Un refroidissement

a) On calcule l'écart à l'air, $\theta - 20$: 180 ; 117 ; 78,2 ; 54,3. Les quotients valent 0,65, environ 0,668 et 0,694 : ils sont proches de 0,67, l'écart suit un modèle exponentiel. Sur θ directement, les quotients (0,685, puis 0,717) sont moins réguliers : c'est bien l'**écart** qu'il faut modéliser.

b) Le coefficient pour 5 min vaut environ 0,67, donc par minute $0,67^{\frac{1}{5}} \approx 0,923$:

$$\theta(t) \approx 20 + 180 \times 0,923^t$$

Contrôle en $t = 10$: l'écart vaut $180 \times 0,923^{10} \approx 80,8$ °C, d'où $\theta \approx 100,8$ °C pour une mesure de 98,2 °C, soit un écart d'environ 2,6 %, acceptable pour un modèle simplifié.

Réponse. a) les quotients des écarts valent environ 0,67 tous les 5 min · b) $\theta(t) \approx 20 + 180 \times 0,923^t$.

Correction de l'exercice 4 - Une durée de vie

a) On calcule les produits xy : $50 \times 240 = 100 \times 120 = 150 \times 80 = 200 \times 60 = 12\,000$. Ils sont constants : la durée de vie est inversement proportionnelle à la vitesse.

b) Le modèle s'écrit $y = \frac{12\,000}{x}$, d'où

$$y(250) = \frac{12\,000}{250} = 48 \text{ min}$$

C'est une extrapolation courte, plausible physiquement : la constante 12 000 représente la longueur totale usinée, en mètres. Contrairement à un modèle affine, ce modèle ne donne jamais de durée négative ; à très grande vitesse, il risque toutefois de sous-estimer l'usure.

Réponse. a) inverse, produits constants à 12 000 · b) $y = \frac{12\,000}{x}$, soit 48 min.

Correction de l'exercice 5 - Comparer deux modèles sur la même série

a) À $x = 100$, la droite donne $-120 + 285 = 165$ min, le modèle inverse 120 min, et la mesure vaut 120 min : la droite se trompe de 45 min, soit 37 %. À $x = 200$, la droite donne 45 min, le modèle inverse 60 min, comme la mesure.

b) À $x = 300$:

$$-1,2 \times 300 + 285 = -75 \text{ min}$$

La droite prévoit une durée **négative**, ce qui est absurde, alors que le modèle inverse donne $\frac{12\,000}{300} = 40$ min, plausible. Seul le modèle inverse respecte la nature de la

grandeur : le contrôle de vraisemblance départage les deux modèles.

Réponse. a) l'affine se trompe de 37 % à $x = 100$ · b) il donne -75 min, absurde : seul le modèle inverse tient.

Correction de l'exercice 6 - Un rendement qui plafonne

Les x doublent à chaque colonne (1, 2, 4, 8) et les y augmentent toujours de 15. Avec $t = \log x$, on obtient 0 ; 0,301 ; 0,602 ; 0,903 : t augmente de $\log 2 \approx 0,301$ à chaque colonne, et y de 15. Le nuage (t ; y) est aligné.

Le coefficient directeur vaut $\frac{15}{0,301} \approx 49,83$, et $y = 40$ pour $t = 0$:

$$y \approx 49,83 \log x + 40$$

Contrôle en $x = 8$: $49,83 \times 0,903 + 40 \approx 85$. Doubler la durée fait toujours gagner 15 points de rendement ; mais le modèle finira par dépasser 100 %.

Réponse. Avec $t = \log x$, le nuage s'aligne : $y \approx 49,83 \log x + 40$.

Correction de l'exercice 7 - Quand le modèle devient impossible

On remplace x par 20 : $\log 20 \approx 1,301$, donc

$$y \approx 49,83 \times 1,301 + 40 \approx 104,8 \%$$

Un rendement supérieur à 100 % est **impossible** : c'est une extrapolation hors du domaine de validité (les mesures vont de 1 à 8 h).

Le modèle atteint 100 % lorsque $49,83 \log x + 40 = 100$, soit $\log x = \frac{60}{49,83} \approx 1,204$ et $x = 10^{1,204} \approx 16$ h : au-delà, il n'a plus de sens. Un modèle réaliste devrait plafonner sous 100 %.

Réponse. Il donne 104,8 %, impossible : le modèle cesse d'être valable vers 16 h.

Correction de l'exercice 8 - Une série à deux régimes

Les différences successives valent 10, 10, 10, puis 5 et 2 : elles sont constantes sur les quatre premiers points, puis chutent brutalement. Un **seul** modèle affine ne convient pas.

On observe deux régimes : jusqu'à $x = 4$, une croissance linéaire, $y = 10x$; ensuite, un net ralentissement, signe probable d'une saturation.

$$y = 10x \quad \text{pour } x \leq 4$$

La bonne pratique consiste à ajuster séparément chaque plage, ou à chercher un modèle qui plafonne. Forcer une droite sur l'ensemble décrirait mal les deux régimes, même si le nuage global paraît assez bien aligné.

Réponse. Non : les différences passent de 10 à 5 puis 2 ; il y a deux régimes et une saturation.

Correction de l'exercice 9 - Une prévision et son incertitude

On remplace x dans l'équation de la droite.

— Pour $x = 30$: $2,4 \times 30 + 18 = 72 + 18 = 90$. La valeur 30 est au milieu de la plage mesurée et les points sont bien alignés : c'est une **interpolation fiable**.

— Pour $x = 120$:

$$2,4 \times 120 + 18 = 288 + 18 = 306$$

La valeur 120 dépasse le double de la borne haute (50) : c'est une **extrapolation**

lointaine, peu fiable, car rien ne dit que la relation reste affine si loin. Un bon alignement sur $[10; 50]$ ne garantit rien en dehors ; il faudrait de nouvelles mesures au-delà de 50.

Réponse. $y(30) = 90$ (interpolation fiable) et $y(120) = 306$ (extrapolation lointaine, peu fiable).

Correction de l'exercice 10 - Deux capteurs à comparer

Les deux capteurs ont la même droite d'étalonnage : même sensibilité (0,05) et même décalage (1,2). Ce qui les distingue, c'est l'écart des points de mesure à la droite.

écart maximal: 0,01 V pour A 0,08 V pour B

Les points du capteur A sont huit fois plus proches de sa droite : une mesure de A, convertie par la droite, est bien plus **précise**. On choisit le capteur **A**.

Le capteur B pourrait suffire pour une mesure indicative, mais pas pour un contrôle exigeant. Deux capteurs de même loi peuvent avoir des précisions très différentes.

Réponse. Le capteur A : même loi, mais ses points sont huit fois plus proches de la droite, donc ses mesures sont plus précises.

Correction de l'exercice 11 - Revenir en arrière

a) Puisque $z = \log y$, on revient à y par la puissance de 10 :

$$y = 10^{2,4771} \times (10^{-0,0212})^x \approx 300 \times 0,9524^x$$

La grandeur baisse d'environ 4,76 % par heure.

b) La demi-vie vérifie $0,9524^x = 0,5$, d'où $x = \frac{\log 0,5}{\log 0,9524} \approx 14,2$ h. On peut aussi diviser directement par le coefficient de l'ajustement, car $\log 0,9524 = -0,0212$:

$$\frac{\log 0,5}{-0,0212} \approx 14,2. \text{ Vérification : } 300 \times 0,9524^{14,2} \approx 150.$$

Réponse. a) $y \approx 300 \times 0,9524^x$ · b) environ 14,2 h.

Correction de l'exercice 12 - Détecter une mesure aberrante

On calcule les quotients de deux valeurs consécutives : $\frac{24}{12} = 2$ et $\frac{48}{24} = 2$, puis $\frac{55}{48} \approx 1,15$ et $\frac{192}{55} \approx 3,49$. La régularité se rompt autour de $x = 4$.

Les trois premiers points suggèrent $y = 6 \times 2^x$. Ce modèle prévoit 96 en $x = 4$, alors qu'on a relevé 55 ; et il prévoit

$$6 \times 2^5 = 192$$

en $x = 5$, exactement la valeur relevée. C'est donc la mesure en $x = 4$ qui est aberrante. On la refait, ou on l'écarte en le signalant, mais on ne la garde pas sans commentaire.

Réponse. La mesure en $x = 4$ (55 au lieu de 96 attendus) est aberrante.

Correction de l'exercice 13 - Un modèle sur données corrigées

Sans la mesure aberrante, il reste les points $x = 1, 2, 3$ et 5. Les quotients consécutifs valent 2 et 2, et de $x = 3$ à $x = 5$, on multiplie par $\frac{192}{48} = 4 = 2^2$: le coefficient par unité vaut bien 2. La valeur en $x = 0$ vaut $\frac{12}{2} = 6$, d'où

$$y = 6 \times 2^x \quad y(8) = 6 \times 256 = 1\,536$$

C'est une extrapolation de trois unités, acceptable si le régime se maintient. Mais un doublement à chaque unité ne dure jamais longtemps : en $x = 15$, on atteindrait 196 608.

Réponse. $y = 6 \times 2^x$, donc $y(8) = 1\,536$.

Correction de l'exercice 14 - Le point moyen après transformation

- a) Le point correspondant a pour ordonnée $y = 10^{2,1} \approx 125,9$.
 b) **Non** : 2,1 est la moyenne des $\log y_i$, pas le logarithme de la moyenne des y_i . Or la moyenne des logarithmes est le logarithme de la **moyenne géométrique** : $10^{2,1}$ est la moyenne géométrique des y .

Par exemple, avec $y_1 = 10$ et $y_2 = 1\,000$, la moyenne arithmétique vaut 505, alors que

$$\frac{\log 10 + \log 1\,000}{2} = 2 \quad \text{et} \quad 10^2 = 100 = \sqrt{10 \times 1\,000}$$

La moyenne géométrique est toujours inférieure ou égale à la moyenne arithmétique ; c'est la moyenne adaptée aux évolutions en pourcentage.

Réponse. a) environ 125,9 · b) non, c'est la moyenne **géométrique** des y .

Correction de l'exercice 15 - Deux transformations possibles

Les deux transformations donnent des nuages presque aussi bien alignés : les données seules ne permettent pas de trancher. Il faut d'autres critères.

- **La nature du phénomène.** Si la grandeur décroît d'un pourcentage constant (refroidissement, décharge), le modèle exponentiel s'impose. Si deux grandeurs ont un produit constant (débit et durée, vitesse et temps), c'est le modèle inverse.
- **Le comportement en extrapolation.** Les deux modèles décroissent vers 0, mais pas au même rythme : ils divergent loin de la plage mesurée.
- **Les résidus.** Si, pour l'un des modèles, les résidus sont tous du même signe sur une plage, ce modèle est mal choisi.

$$y = A \times q^x \quad \text{ou} \quad y = \frac{1}{ax + b}$$

On dit que les deux conviennent statistiquement, et l'on justifie son choix par le contexte.

Réponse. Les données ne suffisent pas : on tranche par la nature physique du phénomène, le comportement en extrapolation et l'examen des résidus.

Correction de l'exercice 16 - Synthèse : l'usure d'un jeu de matrices

- a) Les différences valent 3 ; 4,8 ; 7,7 ; 12,3 : elles augmentent. Les quotients valent 1,6 à chaque fois (aux arrondis près) : le nuage est **exponentiel**.
 b) En comptant les tranches de 10 000 pièces, $x = \frac{n}{10}$, on multiplie par 1,6 à chaque tranche :

$$y = 5 \times 1,6^x$$

soit une hausse de 60 % de l'écart par tranche. **Contrôle** en $n = 40$:

$$5 \times 1,6^4 = 5 \times 6,5536 \approx 32,8 \text{ } \mu\text{m}.$$

- c) On résout $5 \times 1,6^x = 50$, soit $1,6^x = 10$: $x = \frac{\log 10}{\log 1,6} \approx 4,9$. Il faut remplacer le jeu vers 49 000 **pièces**.

d) Le coût de l'outillage vaut $\frac{9\,600}{49} \approx 196$ € par millier de pièces. En tolérant $80\text{ }\mu\text{m}$, $1,6^x = 16$ donne $x \approx 5,9$, soit 59 000 pièces et $\frac{9\,600}{59} \approx 163$ € par millier : environ 17 % d'économie. Cette stratégie n'est rentable que si une dérive de $80\text{ }\mu\text{m}$ reste dans la tolérance de la pièce ; sinon, les rebuts annulent l'économie. Gagner 10 000 pièces coûte $30\text{ }\mu\text{m}$ de dérive : la dérive s'accélère, comme toute exponentielle.

Réponse. a) exponentielle · b) $y = 5 \times 1,6^x$ avec $x = n / 10$, soit +60 % par 10 000 pièces · c) environ 49 000 pièces · d) environ 196 € par millier, contre 163 € en tolérant $80\text{ }\mu\text{m}$ — rentable si la pièce l'accepte.