

Plaquette 2 - Nuage de points, point moyen et ajustement affine

Séries statistiques à deux variables — Terminale STI2D

Corrections détaillées

Boîte à outils

Série à deux variables. Chaque individu porte **deux** mesures, x et y . On les représente par un **nuage de points** $M_i(x_i; y_i)$.

Point moyen.

$$G(\bar{x}; \bar{y}) \quad \text{avec} \quad \bar{x} = \frac{x_1 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + \dots + y_n}{n}$$

La droite d'ajustement passe **toujours** par G : c'est le contrôle le plus rapide d'un résultat de calculatrice.

Droite d'ajustement affine (méthode des moindres carrés) :

$$y = ax + b$$

Coefficient	Sens
a	variation moyenne de y quand x augmente de 1
b	valeur estimée de y pour $x = 0$

Qualité de l'ajustement. Le **résidu** d'un point est l'écart entre la valeur mesurée et la valeur donnée par la droite. La droite des moindres carrés est celle qui rend la plus petite possible la somme

$$S = \sum_i (y_i - (ax_i + b))^2$$

- Résidus petits et de signes alternés : ajustement pertinent.
- Résidus qui dessinent un arc (même signe aux extrémités) : le nuage est incurvé, il faut un autre modèle.

Interpolation et extrapolation.

- **Interpoler** (estimer à l'intérieur de la plage mesurée) : fiable.
- **Extrapoler** (au-delà) : à justifier, et d'autant plus risqué qu'on s'éloigne.

⚠ Un nuage bien aligné ne prouve **aucune** relation de cause à effet : il signale seulement un lien statistique.

Correction de l'exercice 1 - Calculer un point moyen

Le point moyen G a pour coordonnées la moyenne des x et la moyenne des y .

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10}{5} = \frac{30}{5} = 6 \quad \bar{y} = \frac{5 + 9 + 12 + 17 + 22}{5} = \frac{65}{5} = 13$$

Le point moyen est donc $G(6; 13)$.

Contrôle : chaque moyenne est comprise entre la plus petite et la plus grande valeur de sa ligne (6 entre 2 et 10, 13 entre 5 et 22). Réponse : $G(6; 13)$.

Réponse. $G(6; 13)$.

Correction de l'exercice 2 - Vérifier qu'une droite passe par G

La droite d'ajustement par les moindres carrés passe toujours par le point moyen $G(6; 13)$ calculé à l'exercice précédent. On remplace donc x par 6 dans l'équation donnée :

$$2,1 \times 6 + 0,4 = 12,6 + 0,4 = 13 = \bar{y}$$

La droite passe bien par G : le résultat de la calculatrice est **cohérent**. Ce contrôle prend dix secondes et détecte la plupart des erreurs de saisie ; si l'on avait trouvé autre chose que 13, il aurait fallu refaire le calcul.

Réponse. La droite passe bien par $G(6; 13)$: $2,1 \times 6 + 0,4 = 13$.

Correction de l'exercice 3 - Interpréter les coefficients

Le coefficient 3,4 est le coefficient directeur de la droite : quand x augmente de 1, y augmente de 3,4. Chaque heure de fonctionnement supplémentaire consomme donc environ 3,4 **litres** : c'est une consommation horaire, en litres par heure.

Le nombre 18 est l'ordonnée à l'origine : c'est la consommation estimée pour $x = 0$ heure, que l'on peut interpréter comme une consommation de mise en route ou de veille. Mais $x = 0$ est souvent hors de la plage mesurée : cette valeur est une extrapolation, à ne pas surinterpréter.

Par exemple, pour 12 h : $3,4 \times 12 + 18 = 58,8$ L.

Réponse. 3,4 L consommés par heure supplémentaire, 18 L de consommation estimée à l'arrêt.

Correction de l'exercice 4 - Estimer une valeur

On remplace x dans l'équation de la droite.

- Pour 9 h : $3,4 \times 9 + 18 = 30,6 + 18 = 48,6$ L. Comme 9 appartient à la plage mesurée $[4; 16]$, c'est une **interpolation**, fiable.
- Pour 25 h :

$$3,4 \times 25 + 18 = 85 + 18 = 103 \text{ L}$$

Comme 25 est en dehors de la plage mesurée, c'est une **extrapolation** : rien ne garantit que la consommation reste affine sur une aussi longue durée. On donne la valeur **et** sa limite.

Réponse. 48,6 L (interpolation fiable) et 103 L (extrapolation à prendre avec réserve).

Correction de l'exercice 5 - Juger un ajustement par les résidus

On calcule chaque résidu, c'est-à-dire l'écart entre la valeur mesurée et la valeur donnée par la droite :

x	1	2	3	4	5
valeur de la droite	3	5	7	9	11
résidu	0,1	-0,1	0,2	-0,2	0

Les résidus sont petits (au plus 0,2, pour des valeurs de 3 à 11) et leurs signes alternent : les points sont répartis de part et d'autre de la droite, sans courbure. La somme des carrés des résidus vaut

$$0,01 + 0,01 + 0,04 + 0,04 + 0 = 0,10$$

L'ajustement est **bon**. **Contrôle** : la droite passe par le point moyen $G(3; 7)$, car $2 \times 3 + 1 = 7$.

Réponse. Résidus 0,1 ; -0,1 ; 0,2 ; -0,2 ; 0 : petits et de signes alternés, l'ajustement est bon.

Correction de l'exercice 6 - Reconnaître un nuage non aligné

Pour qu'un nuage soit aligné, il faut que y augmente toujours de la même quantité quand x augmente de 1. Les différences valent ici 2, 4, 8 et 16 : elles ne sont pas constantes, les points ne sont **pas** alignés.

En revanche, les quotients valent tous 2 : on multiplie par 2 à chaque pas. Le modèle est exponentiel, $y = 2^x$, et un ajustement affine serait mauvais.

$$\log y = x \log 2$$

En posant $z = \log y$, la relation devient affine : c'est le changement de variable étudié dans la plaquette suivante.

Réponse. Non : les quotients sont constants (2), le nuage est exponentiel, pas aligné.

Correction de l'exercice 7 - Un nuage bien aligné

Les abscisses avancent de 10 en 10. Les différences des y valent 15, 18, 13 et 16 : elles ne sont pas exactement constantes, mais elles restent proches de 15,5 en moyenne. Les points sont **presque** alignés, et un ajustement affine est pertinent.

Le point moyen vaut $\bar{x} = 30$ et $\bar{y} = \frac{371}{5} = 74,2$. Une droite approchée de pente 1,55 passant par G a pour équation

$$y = 1,55x + 27,7$$

Contrôle : $1,55 \times 30 + 27,7 = 74,2$. Réponse : **oui**, la droite passe par $G(30; 74,2)$.

Réponse. Oui : les différences sont proches de 15,5 pour un pas de 10 ; la droite passe par $G(30; 74,2)$.

Correction de l'exercice 8 - Déterminer une droite par deux points

Le coefficient directeur de la droite passant par les deux points est le quotient de la variation des y par celle des x :

$$a = \frac{74 - 18}{25 - 5} = \frac{56}{20} = 2,8$$

On trouve b en écrivant que le premier point est sur la droite : $18 = 2,8 \times 5 + b = 14 + b$, donc $b = 4$. La droite a pour équation $y = 2,8x + 4$.

Contrôle au dernier point : $2,8 \times 25 + 4 = 74$. Cette droite n'est qu'une approximation de la droite des moindres carrés : elle ne passe pas forcément par le point moyen.

Réponse. $y = 2,8x + 4$.

Correction de l'exercice 9 - Comparer une valeur mesurée et estimée

L'estimation donnée par le modèle vaut

$$1,8 \times 20 + 12 = 36 + 12 = 48$$

L'écart entre la mesure et l'estimation vaut $50 - 48 = 2$. Cet écart s'appelle le **résidu** ; il est positif, donc le point mesuré est au-dessus de la droite.

En valeur relative, $\frac{2}{50} = 4\%$: c'est acceptable pour un ajustement. Si tous les résidus

d'une série étaient du même signe, la droite serait mal placée ou le modèle mal choisi.

Réponse. Un écart de 2 (la mesure dépasse l'estimation de 4 %).

Correction de l'exercice 10 - Deux séries à comparer

Les deux études donnent la même droite, mais pas la même dispersion des points autour d'elle. La somme des carrés des résidus mesure cette dispersion : c'est précisément la quantité que la méthode des moindres carrés rend la plus petite possible.

$$S_A = 120 \quad S_B = 6$$

Les points de l'étude B sont vingt fois plus proches de la droite, au sens de cette somme : ses prévisions sont **bien plus fiables**. L'équation d'une droite d'ajustement ne dit rien de la qualité de l'ajustement ; il faut toujours regarder le nuage ou les résidus.

Réponse. Non : les points de B sont beaucoup plus proches de la droite (6 contre 120), ses prévisions sont plus fiables.

Correction de l'exercice 11 - Corrélation n'est pas causalité

Le nuage est presque parfaitement aligné : il existe un fort **lien statistique** entre les deux grandeurs. Mais cela ne prouve pas que les ventilateurs causent des coups de soleil, ce qui serait absurde.

Les deux grandeurs dépendent d'une **troisième** : la saison, ou la température. En été, on achète des ventilateurs **et** on s'expose au soleil ; en hiver, on fait l'inverse.

$$\text{lien statistique} \neq \text{relation de cause à effet}$$

Pour établir une cause, il faudrait une expérience contrôlée ou un mécanisme physique.
Réponse : **non**.

Réponse. Non : une variable cachée (la saison) explique les deux ; un lien statistique n'est pas une relation de cause à effet.

Correction de l'exercice 12 - Un étalonnage

a) Les températures avancent de 20 en 20 et les tensions augmentent à chaque fois de 0,24 V (0,74 – 0,50, 0,98 – 0,74, etc.). Une même variation de x produit toujours la même variation de y : les points sont **exactement** alignés.

b) Le coefficient directeur vaut $\frac{0,24}{20} = 0,012 \text{ V/}^\circ\text{C}$, et la droite passe par (0 ; 0,50) :

$$y = 0,012x + 0,50$$

Contrôle en $x = 80$: $0,96 + 0,50 = 1,46$. Le coefficient 0,012 est la **sensibilité** du capteur, et 0,50 V sa tension à 0 °C.

Réponse. a) les différences valent toutes 0,24 · b) $y = 0,012x + 0,50$.

Correction de l'exercice 13 - Utiliser un étalonnage à l'envers

On cherche la température x pour laquelle la tension vaut 1,10 V :

$$0,012x + 0,50 = 1,10 \Leftrightarrow 0,012x = 0,60 \Leftrightarrow x = \frac{0,60}{0,012} = 50$$

La température vaut 50 °C. **Vérification** : $0,012 \times 50 + 0,50 = 1,10$.

Comme 50 est dans la plage d'étalonnage [0 ; 80], l'estimation est fiable. En général,

$x = \frac{y - 0,50}{0,012}$: c'est la formule que l'on programme dans le calculateur du capteur.

Réponse. 50 °C.

Correction de l'exercice 14 - Une droite décroissante

a) On calcule les deux moyennes :

$$\bar{x} = \frac{60 + 80 + 100 + 120}{4} = 90 \quad \bar{y} = \frac{180 + 140 + 105 + 62}{4} = \frac{487}{4} = 121,75$$

Le point moyen est $G(90; 121,75)$.

b) Quand la vitesse augmente, la durée de vie diminue : la droite d'ajustement sera **décroissante**, de coefficient directeur **négatif**. Une estimation par les points extrêmes donne $\frac{62 - 180}{120 - 60} \approx -1,97$. Les différences successives $(-40, -35, -43)$ sont assez régulières pour qu'un ajustement affine reste acceptable. Physiquement, couper plus vite use l'outil plus vite.

Réponse. a) $G(90; 121,75)$ · b) négatif, car y décroît quand x croît.

Correction de l'exercice 15 - Lire une prévision avec prudence

On remplace x par 300 dans l'équation de la droite :

$$-1,97 \times 300 + 299 = -591 + 299 = -292$$

Le modèle prévoit une durée de vie **négative**, ce qui est absurde. La cause : 300 m/min est très loin de la plage mesurée (60 à 120 m/min), deux fois et demie au-delà de la borne haute. Le modèle affine n'a aucune raison d'y être valable.

Une durée de vie doit décroître vers 0 sans jamais devenir négative : un modèle exponentiel ou inverse serait plus réaliste. Toute extrapolation lointaine doit être confrontée au bon sens physique.

Réponse. Absurde : le modèle donne -292 min, une durée de vie négative ; l'extrapolation est trop lointaine.

Correction de l'exercice 16 - Synthèse : une étude complète

a) $\bar{x} = \frac{1\,500}{5} = 300$ et $\bar{y} = \frac{12 + 19 + 27 + 33 + 41}{5} = \frac{132}{5} = 26,4$, d'où $G(300; 26,4)$.

b) La droite passe par $(100; 12)$ et $(500; 41)$. Son coefficient directeur vaut $\frac{41 - 12}{500 - 100} = \frac{29}{400} = 0,0725$, et $12 = 0,0725 \times 100 + b$ donne $b = 4,75$:

$$y = 0,0725x + 4,75$$

c) En $x = 300$, la droite donne $21,75 + 4,75 = 26,5$, très proche de $\bar{y} = 26,4$: la droite passe presque par G , l'ajustement est bon. Chaque kilogramme supplémentaire consomme 0,0725 kWh, soit 7,25 kWh pour 100 kg ; 4,75 kWh est la consommation à vide, plausible pour une machine.

d) Pour 250 kg, $0,0725 \times 250 + 4,75 \approx 22,9$ kWh : interpolation, fiable. Pour 900 kg, $65,25 + 4,75 = 70$ kWh : extrapolation presque au double de la plage. Il faut l'assortir de réserves : la machine supporte-t-elle 900 kg ? La consommation ne croît-elle pas plus vite en surcharge ?

Réponse. a) $G(300; 26,4)$ · b) $y = 0,0725x + 4,75$ · c) elle donne 26,5 en G , très proche · d) environ 22,9 kWh (fiable) et 70 kWh (extrapolation à justifier).