

Plaquette 4 - Interpoler, extrapoler, critiquer un modèle

Séries statistiques à deux variables — Première STMG

Corrections détaillées

Boîte à outils

Estimer y connaissant x : on remplace x dans $y = ax + b$.

Estimer x connaissant y : on résout l'équation, ou l'inéquation pour un seuil :

$$ax + b = k \Leftrightarrow x = \frac{k - b}{a} \quad (a \neq 0)$$

Interpolation ou extrapolation ?

Valeur de x	Nom	Fiabilité
parmi les valeurs observées	interpolation	bonne
hors des valeurs observées	extrapolation	douteuse

Questions à se poser avant de conclure.

- Le résultat est-il plausible ? (une quantité négative, un pourcentage supérieur à 100 %, une taille de 2,50 m sont absurdes)
- Le nuage a-t-il gardé la même tendance sur toute la période ?
- Une donnée est-elle aberrante (erreur de saisie, événement exceptionnel) ?
- Un lien statistique ne prouve pas que x **cause** y .

Série chronologique. Avec des rangs, l'année A correspond au rang A moins l'année qui précède le rang 1.

Correction de l'exercice 1 - Interpolation ou extrapolation ?

Les données couvrent les rangs 1 à 5. Une valeur de x dans cet intervalle donne une **interpolation**, une valeur en dehors une **extrapolation**.

- 3,5 est entre 1 et 5 : **interpolation**, fiable si le nuage est bien aligné.
- 7 dépasse 5 : **extrapolation** proche (deux ans après la dernière donnée), raisonnable si la tendance se maintient.
- 0 est avant 1 : **extrapolation** vers le passé, d'un an seulement, plutôt raisonnable.
- 2040 correspond au rang 21 : **extrapolation** très lointaine, peu fiable. En vingt ans, le marché immobilier peut changer complètement de tendance.

Plus on s'éloigne des données, moins l'estimation est sûre.

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

Correction de l'exercice 2 - Estimer un loyer

On remplace x dans l'équation de la droite.

a) $y = 14,4 \times 3,5 + 595,8 = 50,4 + 595,8 = 646,2$: environ 646 € (interpolation).

b) 2026 correspond au rang 7 :

$$y = 14,4 \times 7 + 595,8 = 100,8 + 595,8 = 696,6$$

Le loyer est estimé à environ 697 € en 2026 (extrapolation sur deux ans).

c) Le loyer moyen augmente d'environ 14,40 € par an. Sur un loyer d'environ 640 €, cela représente une hausse d'un peu plus de 2 % par an.

Réponse. a) environ 646 € — b) environ 697 € — c) +14,40 € par an.

Correction de l'exercice 3 - Quand le loyer dépassera-t-il 700 euros ?

On résout l'inéquation $14,4x + 595,8 > 700$:

$$14,4x > 104,2 \Rightarrow x > \frac{104,2}{14,4} \approx 7,24$$

Le premier rang entier qui convient est 8. Le rang 1 correspondant à 2020, le rang 8 correspond à $2019 + 8 = 2027$.

Contrôle : au rang 7 (2026), $14,4 \times 7 + 595,8 = 696,6 < 700$; au rang 8 (2027), $14,4 \times 8 + 595,8 = 711 > 700$.

Le loyer moyen devrait dépasser 700 € en 2027, si la tendance observée de 2020 à 2024 se poursuit.

Réponse. En 2027 (rang 8).

Correction de l'exercice 4 - Une extrapolation absurde

a) $6 \times 2 + 75 = 87$, $6 \times 4 + 75 = 99$, $6 \times 6 + 75 = 111$, $6 \times 8 + 75 = 123$ et $6 \times 10 + 75 = 135$: les cinq points sont exactement sur la droite. Entre 2 et 10 ans, la taille augmente de 6 cm par an.

b) À 12 ans : $y = 6 \times 12 + 75 = 147$ cm, ce qui est plausible (extrapolation proche).

À 30 ans :

$$y = 6 \times 30 + 75 = 255 \text{ cm}$$

soit 2,55 m : c'est **absurde**. La croissance s'arrête à la fin de l'adolescence ; le modèle affine n'est valable que pendant l'enfance. Une droite prolonge indéfiniment la tendance observée, ce que la réalité ne fait presque jamais.

Réponse. a) les cinq points vérifient $y = 6x + 75$ — b) 147 cm à 12 ans ; 255 cm à 30 ans, absurde.

Correction de l'exercice 5 - Un taux qui ne peut pas dépasser 100 %

a) Points moyens des deux groupes de trois années :

$$G_1\left(2; \frac{64 + 69 + 71}{3}\right) = (2; 68) \quad G_2\left(5; \frac{77 + 79 + 84}{3}\right) = (5; 80)$$

$a = \frac{80 - 68}{5 - 2} = 4$ et $b = 68 - 4 \times 2 = 60$: la droite de Mayer est $y = 4x + 60$.

b) Au rang 10 : $4 \times 10 + 60 = 100$, tous les foyers seraient équipés. Au rang 11 : $4 \times 11 + 60 = 104$ %, ce qui est **impossible**.

Un pourcentage de foyers équipés ne dépasse jamais 100 %. En réalité, la progression ralentit à l'approche de la saturation : les derniers foyers s'équipent lentement. Le modèle affine ne convient plus quand on approche de 100 %.

Réponse. a) $y = 4x + 60$ — b) 100 % puis 104 %, impossible : le modèle cesse d'être valable.

Correction de l'exercice 6 - Estimer x à partir de y

On cherche x tel que $y = 40$. On résout l'équation :

$$3,75x + 15,5 = 40 \Leftrightarrow 3,75x = 24,5 \Leftrightarrow x = \frac{24,5}{3,75} \approx 6,53$$

Il faudrait dépenser environ 6 530 € en publicité.

Contrôle : $3,75 \times 6,53 + 15,5 \approx 39,99$, soit environ 40 milliers d'euros.

La valeur trouvée est comprise entre 1 et 7, les dépenses observées : il s'agit d'une interpolation, donc d'une estimation raisonnable, à condition que les autres facteurs (saison, prix) ne changent pas.

Réponse. Environ 6 530 € de publicité.

Correction de l'exercice 7 - Comparer une prévision à la réalité

a) 2025 correspond au rang 7 : $y = 7 \times 7 + 113 = 162$, soit une prévision de 162 000 €.

b) L'écart absolu est $158 - 162 = -4$ milliers d'euros : l'entreprise a réalisé 4 000 € de moins que prévu. L'écart relatif rapporte cet écart à la prévision :

$$\frac{158 - 162}{162} = \frac{-4}{162} \approx -0,0247$$

soit environ $-2,5\%$. La prévision était donc assez bonne : l'erreur est faible. Un tel écart peut venir d'un léger ralentissement ; si les écarts négatifs se répètent les années suivantes, il faudra revoir le modèle.

Réponse. a) 162 000 € — b) -4 000 €, soit environ $-2,5\%$.

Correction de l'exercice 8 - Une valeur aberrante

a) Avec la valeur erronée :

$$\bar{y} = \frac{30 + 32 + 35 + 80 + 39 + 41}{6} = \frac{257}{6} \approx 42,83$$

Avec la valeur corrigée : $\bar{y} = \frac{214}{6} \approx 35,67$. Une seule donnée fausse déplace le point moyen d'environ 7 milliers d'euros.

b) Le point aberrant (4 ; 80) est très loin des autres, qui sont presque alignés. Une droite d'ajustement, qui passe par le point moyen, serait tirée vers le haut, et ses estimations seraient trop élevées pour toutes les semaines.

Avant tout ajustement, on regarde le nuage : un point isolé très loin des autres doit être vérifié, et corrigé s'il s'agit d'une erreur.

Réponse. a) environ 42,83 avec l'erreur, 35,67 corrigé — b) la droite serait faussée vers le haut.

Correction de l'exercice 9 - Une croissance qui accélère

a) $y = 3,66 \times 1 - 4,13 = -0,47$: la droite prévoit des ventes **négatives** la première année, alors qu'on en a vendu 200. Le modèle affine décrit mal le début de la série.

b) $y = 3,66 \times 7 - 4,13 = 21,49$, soit environ 2 150 vélos, c'est-à-dire à peu près autant que la sixième année (2 100).

Ce n'est pas vraisemblable : les ventes augmentent de plus en plus vite (+1, +2, +3, +5, puis +8 centaines). Si la tendance continue, la septième année dépasserait nettement 21 centaines. Le nuage dessine une courbe, et un ajustement affine n'est pas adapté.

Réponse. a) $-0,47$: valeur négative, absurde — b) environ 21,5 centaines, sous-estimation probable.

Correction de l'exercice 10 - Corrélation n'est pas causalité

a) **Non.** Un nuage de points aligné montre que les deux grandeurs **varient ensemble** ; il ne dit pas que l'une provoque l'autre.

b) Les deux ventes dépendent d'une troisième grandeur : l'**ensoleillement** (ou la température). Les jours de grand soleil, les vacanciers achètent davantage de glaces **et** prennent davantage de coups de soleil.

Un ajustement affine $y = ax + b$, où x est le nombre de glaces vendues et y le nombre de tubes de crème, permet tout de même d'estimer une grandeur à partir de l'autre (prévoir le stock de crème connaissant les ventes de glaces), mais toute interprétation en termes de cause demande une autre analyse. C'est vrai aussi en gestion : des ventes qui suivent la publicité ne prouvent pas, à elles seules, que la publicité en est la cause.

Réponse. a) non — b) un facteur commun, l'ensoleillement, explique les deux.

Correction de l'exercice 11 - Changer de période d'ajustement

a) Les ventes stagnent autour de 50 les quatre premières années, puis augmentent régulièrement de 6 milliers d'euros par an à partir du rang 4.

b) Les points (5 ; 58) et (8 ; 76) donnent

$$a = \frac{76 - 58}{8 - 5} = \frac{18}{3} = 6 \quad b = 58 - 6 \times 5 = 28$$

soit $y = 6x + 28$. Contrôle : $6 \times 6 + 28 = 64$ et $6 \times 7 + 28 = 70$; les quatre points sont alignés.

c) $y = 6 \times 9 + 28 = 82$. Un ajustement sur les huit années mélangerait la période de stagnation et la période de croissance : la droite de Mayer correspondante a une pente d'environ 4 et prévoit environ 77 au rang 9. Pour prévoir, on ajuste sur la période qui reflète la **tendance actuelle**.

Réponse. b) $y = 6x + 28$ — c) 82 milliers d'euros ; la tendance récente est la plus pertinente.

Correction de l'exercice 12 - Repérer une erreur

L'erreur. L'élève a remplacé x par l'**année** 2025 :

$$7 \times 2025 + 113 = 14\,288$$

Or la droite a été calculée avec les **rangs** (1 pour 2019, 2 pour 2020...). Un chiffre d'affaires multiplié par cent en un an aurait d'ailleurs dû l'alerter.

La correction. Le rang de 2025 est $2025 - 2018 = 7$:

$$y = 7 \times 7 + 113 = 162$$

Le chiffre d'affaires estimé pour 2025 est d'environ 162 000 €, cohérent avec les 155 000 € de 2024.

Réponse. Il faut utiliser le rang 7 : environ 162 000 €.

Correction de l'exercice 13 - Lire un résultat de tableur

a) La pente est le coefficient directeur a et l'ordonnée à l'origine est b :

$$y = 14,4x + 595,8$$

b) L'ordonnée à l'origine est la valeur de y pour $x = 0$, c'est-à-dire pour le rang 0, l'année 2019. Le modèle estime le loyer moyen de 2019 à environ 595,80 €.

Ce n'est **pas** une valeur observée : les données commencent en 2020. C'est une extrapolation d'un an vers le passé, plausible mais non mesurée. Ce nombre sert surtout à écrire l'équation ; on évite de l'interpréter quand $x = 0$ n'a pas de sens (par exemple une surface de magasin nulle).

Réponse. a) $y = 14,4x + 595,8$ — b) loyer estimé de 2019, non observé.

Correction de l'exercice 14 - Deux prévisions, deux fiabilités

Rang 4,5 :

$$y = 23,66 \times 4,5 + 397,2 = 106,47 + 397,2 = 503,67$$

soit environ 504 adhérents. C'est une **interpolation** : le rang 4,5 est entre les données, et l'estimation est fiable si le nuage est bien aligné.

Rang 8 : $y = 23,66 \times 8 + 397,2 = 189,28 + 397,2 = 586,48$, soit environ 586 adhérents. C'est une **extrapolation** à deux ans de la dernière donnée : elle suppose que la salle continue de gagner environ 24 adhérents par an.

La première estimation est la plus fiable. On arrondit à l'unité, car on compte des personnes.

Réponse. Environ 504 adhérents (interpolation, fiable) et 586 en 2026 (extrapolation, moins sûre).

Correction de l'exercice 15 - Synthèse : la capacité d'une salle de sport

a) $\bar{x} = \frac{21}{6} = 3,5$ et $\bar{y} = \frac{420 + 446 + 468 + 492 + 514 + 540}{6} = \frac{2880}{6} = 480$, donc $G(3,5 ; 480)$.

b) $23,66 \times 3,5 + 397,2 = 82,81 + 397,2 = 480,01$: on retrouve 480 aux arrondis près.

c) On résout $23,66x + 397,2 > 600$:

$$x > \frac{202,8}{23,66} \approx 8,57$$

Le premier rang qui convient est 9, soit l'année $2018 + 9 = 2027$ ($23,66 \times 9 + 397,2 \approx 610$).

d) La salle ne peut pas accueillir plus de 600 adhérents : au-delà, elle refusera des inscriptions ou devra s'agrandir. Le modèle affine ignore cette contrainte ; il indique seulement **quand** le problème se posera, ce qui est précisément utile pour décider d'un agrandissement.

Réponse. a) $G(3,5 ; 480)$ — b) 480,01 — c) en 2027 — d) la capacité limite la croissance réelle.