

Corrigé et barème – Statistiques — série double et régression

Nuage de points, point moyen, droite des moindres carrés et coefficient de corrélation linéaire
(programme Maths complémentaires Terminale, BO 2019)

Durée : 120 min – Calculatrice interdite – Sur 20 points · Maths complémentaires (option Tle) Terminale

Exercice 1 – Nuage, point moyen et ajustement

(4 pts)

- $\bar{x} = \frac{45}{5} = 9$ et $\bar{y} = \frac{171}{5} = 34,2$, donc $G(9; 34,2)$. (1)
- $\sum (x_i - \bar{x})(y_i - \bar{y}) = 180$ et $\sum (x_i - \bar{x})^2 = 90$, d'où $a = \frac{180}{90} = 2$ puis $b = 34,2 - 2 \times 9 = 16,2$: $y = 2x + 16,2$ (contrôle : la droite passe par G). (1)
- $\sum (y_i - \bar{y})^2 = 362,8$, donc $r = \frac{180}{\sqrt{90 \times 362,8}} = \frac{180}{180,70} \approx 0,996$: les points sont presque alignés, l'ajustement affine est justifié. (1)
- $y = 2 \times 20 + 16,2 = 56,2$ g. Comme $20 \notin [3; 15]$, c'est une extrapolation : l'estimation n'est valable que si la tendance se poursuit. (1)

Repère — Vérifier que $a\bar{x} + b = \bar{y}$ prend cinq secondes et détecte presque toutes les erreurs de calcul sur a ou b .

Exercice 2 – Lecture, résolution et limites d'un modèle

(4 pts)

- 2026 correspond à $x = 10$: $y = -0,48 \times 10 + 31,7 = 26,9$, soit environ 26,9 %. Comme $10 \notin [0; 9]$, c'est une extrapolation, proche de la plage observée donc plausible. (1)
- Chaque année écoulée fait perdre, selon le modèle, environ 0,48 point de pourcentage de foyers équipés. (1)
- $-0,48x + 31,7 < 25$ donne $-0,48x < -6,7$, soit $x > \frac{6,7}{0,48} \approx 13,96$: le premier entier est $x = 14$, c'est-à-dire l'année 2030. (1)
- $-0,48x + 31,7 < 0$ équivaut à $x > \frac{31,7}{0,48} \approx 66,04$, donc à partir de $x = 67$, soit 2083. Un pourcentage négatif étant impossible, le modèle affine ne peut pas être utilisé loin de la plage des données. (1)

Attention — Diviser une inégalité par un nombre négatif change le sens de l'inégalité : ici $-0,48x < -6,7$ donne bien $x > 13,96$.

Exercice 3 – Changement de variable logarithmique

(4 pts)

- $\ln 480 \approx 6,174$; $\ln 350 \approx 5,858$; $\ln 262 \approx 5,568$; $\ln 190 \approx 5,247$; $\ln 140 \approx 4,942$. (1)
- La calculatrice donne $z = -0,308t + 6,480$ avec $r \approx -1,000$ (précisément $-0,9999$) : les points $(t_i; z_i)$ sont quasiment alignés. (1)
- $\ln y = -0,308t + 6,480$ donne $y = e^{6,480} \times e^{-0,308t}$, soit $y \approx 652 e^{-0,308t}$ becquerels. (1)
- $y \approx 652 e^{-0,308 \times 8} = 652 e^{-2,460} \approx 652 \times 0,0854 \approx 55,7$ Bq. Comme $8 \notin [1; 5]$, c'est une extrapolation, ici raisonnable car la décroissance radioactive est effectivement exponentielle. (1)

Erreur fréquente — $e^{\alpha t + \beta}$ vaut $e^{\beta} \times e^{\alpha t}$, et non $e^{\alpha t} + e^{\beta}$: l'exponentielle transforme une somme en produit.

Exercice 4 – Corrélation, résidu et causalité*(4 pts)*

- a) $\text{Cov}(x, y) = 34,2 > 0$: le coefficient directeur a est donc positif et le nuage est croissant (les collègues les mieux équipés ont en moyenne de meilleurs scores). (1)
- b) $a = \frac{34,2}{27,5} \approx 1,244$ et $b = \bar{y} - a\bar{x} = 52,5 - 1,243636 \times 18 \approx 52,5 - 22,39 \approx 30,11$. (1)
- c) Valeur prédite : $1,243636 \times 22 + 30,1145 \approx 57,47$. Résidu : $58,4 - 57,47 \approx 0,93 > 0$, donc ce collègue est au-dessus de la droite : il réussit un peu mieux que ne le prévoit le modèle. (1)
- d) Il confond corrélation et causalité. Une variable cachée — le niveau de ressources des communes — agit à la fois sur l'équipement informatique et sur les résultats scolaires. Les données décrivent une association observée entre collègues, elles ne permettent pas de prévoir l'effet d'une action sur un collègue donné : seule une expérimentation contrôlée le pourrait. (1)

Le point clé — Une droite de régression décrit ce que l'on a observé ; elle ne dit jamais ce qui se passerait si l'on modifiait volontairement x .

Exercice 5 – Un relevé qui détonne*(4 pts)*

- a) $\bar{x} = \frac{61}{6} \approx 10,167$ et $\bar{y} = \frac{113}{6} \approx 18,833$; on obtient $\sum(x_i - \bar{x})(y_i - \bar{y}) \approx 21,17$, $\sum(x_i - \bar{x})^2 \approx 98,83$ et $\sum(y_i - \bar{y})^2 \approx 470,83$, d'où $r \approx \frac{21,17}{215,71} \approx 0,098$. Cette valeur très faible ne signifie pas qu'il n'y a pas de tendance : elle est causée par le dernier relevé, isolé. (1)
- b) Sur les cinq premiers relevés : $\bar{x} = \frac{45}{5} = 9$ et $\bar{y} = \frac{108}{5} = 21,6$, donc $G(9; 21,6)$; $\sum(x_i - \bar{x})(y_i - \bar{y}) = 118$ et $\sum(x_i - \bar{x})^2 = 58$, d'où $a = \frac{118}{58} \approx 2,03$ et $b = 21,6 - 2,0345 \times 9 \approx 3,29$: $y = 2,03x + 3,29$ (et $r \approx 0,998$). (1)
- c) Valeur prédite en $x = 16$: $2,0345 \times 16 + 3,2897 \approx 35,84$ cm. Résidu : $5 - 35,84 = -30,84$ cm, énorme en valeur absolue devant les autres écarts : le relevé est bien aberrant. (1)
- d) On ne peut l'écarter que si une cause extérieure aux données le justifie (plant cassé, erreur de saisie, mauvaise mesure). Le compte rendu doit signaler le relevé, indiquer la raison de son retrait et donner les deux résultats : rien ne doit être supprimé silencieusement. (1)

À retenir — Un r médiocre n'est pas toujours le signe d'une absence de tendance : il faut regarder le nuage avant de conclure.