Chapitre 7 — Statistiques à deux variables

Terminale Technologique • Tronc commun

Table des matières

PROGRAMME (BO — TERMINALE TECHNOLOGIQUE)

Contenus : Série statistique à deux variables quantitatives. Nuage de points associé. Point moyen. Ajustement affine : droite des moindres carrés, méthode des points moyens. Interpolation et extrapolation. Changement de variable pour se ramener à un ajustement affine.

Démonstrations : Aucune démonstration exigible.

Capacités : Représenter un nuage de points dans un repère. Calculer les coordonnées du point moyen. Déterminer une droite d'ajustement (calculatrice ou méthode des points moyens). Utiliser un ajustement pour interpoler ou extrapoler. Effectuer un changement de variable donné et revenir au modèle initial. Exercer un esprit critique sur la pertinence d'un ajustement.

Tout le cours

Activités

Objectif : représenter un nuage de points et calculer un point moyen.

Une entreprise a relevé sur cinq années son budget publicitaire x (en centaines d'euros) et son volume de ventes y (en milliers d'unités).

x_i	10	15	20	25	30
y_i	22	28	35	40	48

1. Représenter le nuage de points $(x_i\,;\,y_i)$ dans un repère.
2. Le nuage semble-t-il aligné ?
3. Calculer les coordonnées du point moyen $G(\overline{x}\,;\,\overline{y})$.
4. Placer G dans le repère. Est-il bien au centre du nuage ?

Correction. 1. Les points sont placés dans un repère.
2. Oui, les points sont approximativement alignés.
3. $\overline{x} = \frac{10+15+20+25+30}{5} = 20$ et $\overline{y} = \frac{22+28+35+40+48}{5} = 34{,}6$. Donc $G(20\,;\,34{,}6)$.
4. G est bien au centre du nuage.

Objectif : utiliser une droite d'ajustement pour faire des estimations.

On reprend les données de l'activité 1. À l'aide de la calculatrice, on obtient la droite de régression $y = 1{,}3x + 8{,}6$.

1. Tracer cette droite dans le repère précédent.
2. Vérifier que la droite passe par G.
3. Estimer le volume de ventes pour un budget de 35 centaines d'euros. S'agit-il d'une interpolation ou d'une extrapolation ?
4. Estimer le budget nécessaire pour vendre 50 milliers d'unités.

Correction. 1. La droite passe par les points $(0\,;\,8{,}6)$ et $(20\,;\,34{,}6)$.
2. $y(20) = 1{,}3 \times 20 + 8{,}6 = 34{,}6 = \overline{y}$. Oui, G est sur la droite.
3. $y(35) = 1{,}3 \times 35 + 8{,}6 = 54{,}1$. Extrapolation (35 est hors de la plage $[10\,;\,30]$).
4. $50 = 1{,}3x + 8{,}6 \Rightarrow x = \frac{50-8{,}6}{1{,}3} \approx 31{,}8$. Budget d'environ 3 180 euros.

Objectif : transformer un nuage courbé en nuage aligné.

x_i	1	2	3	4	5
y_i	20	40	80	160	320

1. Représenter le nuage de points. L'ajustement affine semble-t-il pertinent ?
2. On pose $z_i = \log(y_i)$. Calculer les valeurs de z_i (arrondir au centième).
3. Représenter le nuage $(x_i\,;\,z_i)$. Ce nuage est-il plus aligné ?
4. Déterminer la droite de régression $z = ax + b$ à la calculatrice.
5. En déduire une relation entre y et x.

Correction. 1. Le nuage est courbé, l'ajustement affine n'est pas pertinent.
2. $z_1 = \log(20) \approx 1{,}30$; $z_2 \approx 1{,}60$; $z_3 \approx 1{,}90$; $z_4 \approx 2{,}20$; $z_5 \approx 2{,}51$.
3. Le nuage $(x_i\,;\,z_i)$ est presque aligné.
4. $z \approx 0{,}30x + 1{,}00$.
5. $\log(y) = 0{,}30x + 1{,}00$ donc $y = 10^{0{,}30x+1} = 10 \times 10^{0{,}30x} \approx 10 \times 2^x$.

1 Série statistique à deux variables

Définition. Une **série statistique à deux variables quantitatives** est constituée d'une liste de couples :

$$(x_1\,;\,y_1),\ (x_2\,;\,y_2),\ \dots,\ (x_n\,;\,y_n).$$

Chaque couple associe une valeur du caractère x et une valeur du caractère y.

Définition. Le **nuage de points** associé à la série est l'ensemble des points $M_i(x_i\,;\,y_i)$ représentés dans un repère orthogonal.

Définition. Le **point moyen** G du nuage a pour coordonnées :

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

Le point moyen se situe au « centre » du nuage. On le place toujours dans le repère pour vérifier la cohérence.

Exemple. Calculer le point moyen de la série (1 ; 4), (3 ; 6), (5 ; 8).

Correction. $\overline{x} = \frac{1+3+5}{3} = 3$ et $\overline{y} = \frac{4+6+8}{3} = 6$. Donc $G(3\,;\,6)$.

2 Ajustement affine

Définition. Lorsque les points du nuage semblent presque **alignés**, on modélise la tendance par une droite d'équation :

$$\boxed{y = ax + b}$$

appelée **droite d'ajustement affine**.

Droite des moindres carrés. La **droite de régression** de y en x est la droite $y = ax + b$ qui minimise la somme des carrés des écarts verticaux entre les points observés et la droite.
Les coefficients a et b sont fournis par la **calculatrice** ou un tableur.

Propriété. La droite de régression **passe toujours par le point moyen** $G(\bar{x} ; \bar{y})$.

1. Partager le nuage en **deux groupes** de taille à peu près égale.
2. Calculer le point moyen G_1 du premier groupe et G_2 du second.
3. La droite (G_1G_2) est une droite d'ajustement du nuage.

TI : STAT → EDIT (saisir $L1$, $L2$) → STAT → CALC → RegLin(ax+b).
Casio : STAT → saisir List1, List2 → CALC → SET → REG → X → aX+b.

Exemple. Série : (2 ; 5), (4 ; 9), (6 ; 14), (8 ; 18).

Correction. $\bar{x} = \frac{2+4+6+8}{4} = 5$ et $\bar{y} = \frac{5+9+14+18}{4} = 11{,}5$. Donc $G(5 ; 11{,}5)$.
À la calculatrice : $y \approx 2{,}15x + 0{,}75$.
Vérification : $2{,}15 \times 5 + 0{,}75 = 11{,}5 = \bar{y}$. La droite passe bien par G.

3 Interpolation et extrapolation

Définition.
- On parle d'**interpolation** lorsque l'on estime une valeur **à l'intérieur** de la plage des données observées.
- On parle d'**extrapolation** lorsque l'on estime une valeur **à l'extérieur** de la plage des données observées.

Une **interpolation** est en général plus fiable qu'une extrapolation.
Une **extrapolation** prolonge une tendance qui peut changer en dehors du domaine étudié. Il faut rester prudent.

Exemple. Des données sont relevées pour x entre 1 et 10. On utilise la droite $y = 3x + 2$.

Correction. Estimer y pour $x = 6$: $y = 3 \times 6 + 2 = 20$. C'est une **interpolation** ($6 \in [1\,;\,10]$).
Estimer y pour $x = 15$: $y = 3 \times 15 + 2 = 47$. C'est une **extrapolation** ($15 \notin [1\,;\,10]$).

4 Changement de variable

Méthode. Lorsque le nuage de points n'est pas aligné, on peut effectuer un **changement de variable** (donné dans l'énoncé) pour se ramener à un ajustement affine :

$$z = \log(y), \qquad u = \frac{1}{t}, \qquad z = \sqrt{x}, \quad \text{etc.}$$

Retour au modèle initial.
Cas 1 : $z = \log(y)$ et $z = ax + b \Rightarrow \log(y) = ax + b \Rightarrow \boxed{y = 10^{ax+b} = 10^b \cdot 10^{ax}}$.

Cas 2 : $u = \frac{1}{t}$ et $u = ax + b \Rightarrow \boxed{t = \frac{1}{ax+b}}$.

Cas 3 : $z = \sqrt{x}$ et $y = az + b \Rightarrow \boxed{y = a\sqrt{x} + b}$.

1. Observer le nuage : aligné ou courbé ?
2. Si courbé, effectuer le changement de variable donné.
3. Représenter le nouveau nuage $(x_i ; z_i)$.
4. Réaliser l'ajustement affine sur les nouvelles variables.
5. Revenir au modèle initial en remplaçant z.
6. Distinguer interpolation et extrapolation.

Exemple. On a $z = \log(y)$ et la calculatrice donne $z = 0{,}02x + 1{,}5$. Exprimer y en fonction de x.

Correction. $\log(y) = 0{,}02x + 1{,}5$ donc $y = 10^{0{,}02x+1{,}5} = 10^{1{,}5} \times 10^{0{,}02x} \approx 31{,}6 \times 10^{0{,}02x}$.

5 Bilan

Notion	Résultat
Série à deux variables	Liste de couples $(x_i\,;\,y_i)$
Nuage de points	Points $M_i(x_i\,;\,y_i)$ dans un repère
Point moyen	$G(\overline{x}\,;\,\overline{y})$ avec $\overline{x} = \frac{\sum x_i}{n}$, $\overline{y} = \frac{\sum y_i}{n}$
Droite d'ajustement	$y = ax + b$ (moindres carrés ou points moyens)
Propriété	La droite de régression passe par G
Interpolation	Estimation à l'intérieur des données
Extrapolation	Estimation à l'extérieur des données
Changement de variable	$z = \log(y)$, $u = 1/t$, $z = \sqrt{x}$, etc.
Retour au modèle	$\log(y) = ax + b \Rightarrow y = 10^{ax+b}$

6 Exercice de synthèse

Le tableau suivant donne le chiffre d'affaires y_i (en milliers d'euros) d'une entreprise en fonction de l'année x_i (rang de l'année, $x_i = 0$ pour 2015).

x_i	0	1	2	3	4	5
y_i	55	58	64	85	105	112

1. Représenter le nuage de points dans un repère.

2. Calculer les coordonnées du point moyen G.

3. À l'aide de la calculatrice, déterminer la droite de régression $y = ax + b$ (arrondir au dixième).

4. Tracer cette droite. Vérifier qu'elle passe par G.

5. Estimer le CA en 2021 ($x = 6$). Est-ce une interpolation ou une extrapolation ?

6. Le nuage semblant légèrement courbé, on pose $z_i = \log(y_i)$. Calculer les z_i.

7. Déterminer la droite de régression $z = \alpha x + \beta$.

8. En déduire une expression de y en fonction de x.

9. Comparer les deux prévisions pour 2021.

Correction. 2. $\overline{x} = \frac{0+1+2+3+4+5}{6} = 2{,}5$ et $\overline{y} = \frac{55+58+64+85+105+112}{6} \approx 79{,}8$. $G(2{,}5\,;\,79{,}8)$.

3. À la calculatrice : $y \approx 12{,}8x + 47{,}9$.

4. $12{,}8 \times 2{,}5 + 47{,}9 = 79{,}9 \approx \overline{y}$. OK.

5. $y(6) = 12{,}8 \times 6 + 47{,}9 = 124{,}7$. Extrapolation car $6 \notin [0\,;\,5]$.

6. $z_0 = \log(55) \approx 1{,}74$; $z_1 \approx 1{,}76$; $z_2 \approx 1{,}81$; $z_3 \approx 1{,}93$; $z_4 \approx 2{,}02$; $z_5 \approx 2{,}05$.

7. $z \approx 0{,}066x + 1{,}71$.

8. $\log(y) = 0{,}066x + 1{,}71$ donc $y = 10^{0{,}066x+1{,}71} \approx 51{,}3 \times 10^{0{,}066x}$.

9. Ajustement affine : $y(6) \approx 124{,}7$. Ajustement exponentiel : $y(6) \approx 51{,}3 \times 10^{0{,}396} \approx 127{,}5$. Les prévisions sont proches.