DS Blanc n°1

Statistiques à deux variables

Terminale Techno — Chapitre 7

Exercice 1 — Automatismes (15 min — 3 pts) [Correction]

Calculer les statistiques simples :

1. Moyenne de x : (1 ; 2 ; 3 ; 4 ; 5)
2. Variance de x
3. Écart-type de x

Exercice 2 — Point moyen (15 min — 3 pts) [Correction]

On considère la série double :

x	1	2	3	4
y	3	5	7	9

1. Calculer $\overline{x}$ et $\overline{y}$.
2. Placer le point moyen $G(\overline{x}\,;\,\overline{y})$ sur un graphique.
3. Interpréter le point moyen dans le contexte.

Exercice 3 — Régression (20 min — 5 pts) [Correction]

À partir de la série précédente :

1. Calculer la covariance $\text{Cov}(x,\, y)$.
2. Déterminer la droite de régression $y = ax + b$ en utilisant $a = \dfrac{\text{Cov}(x,\, y)}{V(x)}$.
3. Vérifier que cette droite passe par G.

Exercice 4 — Estimation (15 min — 3 pts) [Correction]

Utiliser la droite de régression trouvée :

1. Estimer la valeur de y pour $x = 5$.
2. Estimer la valeur de y pour $x = 0$.
3. Discuter des limites de l'extrapolation.

Exercice 5 — Changement variable (20 min — 4 pts) [Correction]

On pose $z = \ln(y)$. La nouvelle série devient :

x	1	2	3	4
z	$\ln(3)$	$\ln(5)$	$\ln(7)$	$\ln(9)$

1. Calculer $\overline{x}$ et $\overline{z}$.
2. Déterminer la droite $z = cx + d$.
3. En déduire l'expression de y en fonction de x.

Exercice 6 — Problème (10 min — 2 pts) [Correction]

Un organisme observe le nombre d'adhésions à un club selon le prix d'inscription. Justifier le choix d'un modèle linéaire ou exponentiel selon la tendance observée.

Corrigé — DS Blanc n°1

Statistiques à deux variables

Terminale Techno — Chapitre 7

Correction 1 — Automatismes () [Énoncé]

Soit la série x : $(1\,;2\,;3\,;4\,;5)$

1. Moyenne : $\overline{x} = \dfrac{1+2+3+4+5}{5} = \dfrac{15}{5} = 3$
2. Variance : $V(x) = \dfrac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5}$

$$V(x) = \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$$

3. Écart-type : $\sigma(x) = \sqrt{V(x)} = \sqrt{2} \approx 1{,}41$

Correction 2 — Point moyen () [Énoncé]

Série double :

x	1	2	3	4
y	3	5	7	9

1. Calcul des moyennes :

$$\overline{x} = \frac{1+2+3+4}{4} = \frac{10}{4} = 2{,}5$$
$$\overline{y} = \frac{3+5+7+9}{4} = \frac{24}{4} = 6$$

2. Le point moyen est $G(2{,}5\,;6)$.
3. Le point moyen représente le centre de gravité du nuage de points. C'est le barycentre autour duquel les points sont distribués.

Correction 3 — Régression () [Énoncé]

1. Covariance — on calcule d'abord les produits xy :

x	1	2	3	4
y	3	5	7	9
xy	3	10	21	36

Moyenne des produits : $\overline{xy} = \dfrac{3+10+21+36}{4} = \dfrac{70}{4} = 17{,}5$

$$\text{Cov}(x,\ y) = \overline{xy} - \overline{x}\cdot\overline{y} = 17{,}5 - 2{,}5\times 6 = 17{,}5 - 15 = 2{,}5$$

2. Variance de x :

$$V(x) = \frac{(1-2{,}5)^2+(2-2{,}5)^2+(3-2{,}5)^2+(4-2{,}5)^2}{4} = \frac{2{,}25+0{,}25+0{,}25+2{,}25}{4} = \frac{5}{4} = 1{,}25$$

Coefficient directeur :

$$a = \frac{\text{Cov}(x,\ y)}{V(x)} = \frac{2{,}5}{1{,}25} = 2$$

Ordonnée à l'origine :

$$b = \overline{y} - a\cdot\overline{x} = 6 - 2\times 2{,}5 = 6 - 5 = 1$$

Droite : $y = 2x + 1$

3. Vérification : $G(2{,}5\,;6)$ doit satisfaire $y = 2x + 1$

$$y = 2\times 2{,}5 + 1 = 5 + 1 = 6$$

Correction 4 — Estimation () [Énoncé]

En utilisant la droite $y = 2x + 1$:

1. Pour $x = 5$: $y = 2 \times 5 + 1 = 11$
2. Pour $x = 0$: $y = 2 \times 0 + 1 = 1$
3. Limites de l'extrapolation :
   — L'extrapolation au-delà de la plage observée $[1\,;\,4]$ est risquée car la relation linéaire observée peut ne pas persister.
   — La prédiction pour $x = 5$ sort du domaine d'observation.
   — La prédiction pour $x = 0$ est bien éloignée des données collectées.
   — La qualité de l'ajustement dépend du coefficient de corrélation.

Correction 5 — Changement variable () [Énoncé]

Avec $z = \ln(y)$:

1. Calcul des moyennes :
$$\overline{x} = 2{,}5$$
$$\overline{z} = \frac{\ln(3) + \ln(5) + \ln(7) + \ln(9)}{4} \approx \frac{1{,}099 + 1{,}609 + 1{,}946 + 2{,}197}{4} \approx 1{,}713$$
2. Déterminer la droite $z = cx + d$ — calculer $\mathrm{Cov}(x, z)$ et $V(x)$:
$$c = \frac{\mathrm{Cov}(x, z)}{V(x)} \approx \frac{0{,}4}{1{,}25} \approx 0{,}32$$
$$d = \overline{z} - c \cdot \overline{x} \approx 1{,}713 - 0{,}32 \times 2{,}5 \approx 0{,}913$$
   Droite : $z \approx 0{,}32x + 0{,}91$
3. En déduire y : puisque $z = \ln(y)$:
$$\ln(y) \approx 0{,}32x + 0{,}91$$
$$y \approx e^{0{,}32x+0{,}91} = e^{0{,}91} \cdot e^{0{,}32x} \approx 2{,}48 \cdot e^{0{,}32x}$$

Correction 6 — Problème () [Énoncé]

Pour justifier le choix du modèle :

1. Si le nuage de points s'aligne approximativement sur une droite, un modèle linéaire $y = ax + b$ est approprié.
2. Si le nuage suit une courbe exponentielle (croissance ou décroissance rapide), on utilisera un modèle exponentiel $y = Ae^{kx}$ après transformation logarithmique.
3. Dans ce contexte (adhésions vs prix), on s'attend généralement à une relation décroissante :
   — Linéaire : réduction uniforme du nombre d'adhésions.
   — Exponentielle : baisse accélérée des adhésions quand le prix augmente.
4. Le choix dépend de l'ajustement observé : calculer le coefficient de corrélation pour le modèle linéaire et l'ajustement exponentiel, puis choisir celui avec le meilleur $|r|$.