DS Blanc n°2

Exercice 1 — QCM **(15 min — 4 pts)** [Correction]

Pour chaque question, sélectionner la bonne réponse.

1. Le point moyen G d'un nuage de points :
   - A. Est toujours sur la droite de régression
   - B. Peut ne pas être dans le nuage
   - C. Minimise la somme des écarts
2. Si $\text{Cov}(x, y) < 0$, alors :
   - A. Les variables sont indépendantes
   - B. La droite de régression est croissante
   - C. La droite de régression est décroissante
3. La variance mesure :
   - A. La dispersion autour de la moyenne
   - B. L'écart entre deux valeurs
   - C. La tendance générale
4. L'écart-type est :
   - A. Le carré de la variance
   - B. La racine carrée de la variance
   - C. La variance divisée par la moyenne

Exercice 2 — Points moyens **(15 min — 3 pts)** [Correction]

Données : $(1, 2), (2, 4), (3, 6), (4, 8), (5, 10)$

1. Calculer $\bar{x}$ et $\bar{y}$
2. Justifier que le point moyen G est le centre de gravité du nuage
3. Tracer le nuage et placer G

Exercice 3 — Étude complète **(20 min — 5 pts)** [Correction]

Série : $(0, 1), (1, 3), (2, 5), (3, 7), (4, 9)$

1. Calculer $\bar{x}$, $\bar{y}$, $V(x)$, $V(y)$
2. Déterminer $\text{Cov}(x, y)$ et le coefficient de corrélation r
3. Écrire l'équation de la droite de régression
4. Interpréter le coefficient de corrélation

Exercice 4 — Changement log **(20 min — 4 pts)** [Correction]

On observe une croissance exponentielle. Données brutes :

t	0	1	2	3	4
N	10	20	40	80	160

1. Calculer $y = \ln(N)$ pour chaque valeur
2. Justifier que le changement de variable linéarise les données
3. Déterminer la droite $y = at + b$
4. En déduire N en fonction de t

Exercice 5 — Problème concret (20 min — 4 pts) [Correction]

Une entreprise analyse ses dépenses en publicité et son chiffre d'affaires. Le responsable marketing propose un modèle exponentiel. Expliquer pourquoi cela pourrait être justifié et comment le valider.

Correction 1 — QCM () [Énoncé]

1. A (Oui, le point moyen se situe toujours sur la droite de régression par construction)
2. C (Covariance négative implique une relation décroissante, donc la pente est négative)
3. A (La variance mesure la dispersion des données autour de la moyenne)
4. B (L'écart-type $\sigma = \sqrt{V}$ est par définition la racine carrée de la variance)

Correction 2 — Points moyens () [Énoncé]

Données : $(1,2),(2,4),(3,6),(4,8),(5,10)$

1. Calcul des moyennes :
$$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$
$$\bar{y} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$$
2. Le point moyen $G(3,6)$ est le centre de gravité car il minimise la somme des carrés des distances (dans le sens des moindres carrés). C'est le point tel que le nuage est équilibré autour de lui.
3. Le graphique montre 5 points alignés sur la droite $y = 2x$ avec le point G au centre.

Correction 3 — Étude complète () [Énoncé]

Série : $(0,1),(1,3),(2,5),(3,7),(4,9)$

1. Moyennes :
$$\bar{x} = \frac{0+1+2+3+4}{5} = 2$$
$$\bar{y} = \frac{1+3+5+7+9}{5} = 5$$
Variances :
$$V(x) = \frac{(0-2)^2+(1-2)^2+(2-2)^2+(3-2)^2+(4-2)^2}{5} = \frac{4+1+0+1+4}{5} = 2$$
$$V(y) = \frac{(1-5)^2+(3-5)^2+(5-5)^2+(7-5)^2+(9-5)^2}{5} = \frac{16+4+0+4+16}{5} = 8$$
2. Covariance : Calcul des produits xy :

xy	0	3	10	21	36

$$\overline{xy} = \frac{0+3+10+21+36}{5} = 14$$
$$\mathrm{Cov}(x,y) = 14 - 2\times 5 = 14 - 10 = 4$$
Coefficient de corrélation :
$$r = \frac{\mathrm{Cov}(x,y)}{\sigma(x)\cdot\sigma(y)} = \frac{4}{\sqrt{2}\cdot\sqrt{8}} = \frac{4}{4} = 1$$
3. Droite de régression : $a = \frac{\mathrm{Cov}(x,y)}{V(x)} = \frac{4}{2} = 2$
$$b = \bar{y} - a\cdot\bar{x} = 5 - 2\times 2 = 1$$
Équation : $y = 2x + 1$
4. Interprétation : $r = 1$ indique une corrélation linéaire parfaite. Les points sont exactement alignés sur la droite. La relation entre x et y est déterministe (pas de dispersion autour de la droite).

Correction 4 — Changement log () [Énoncé]

1. Calcul de $y = \ln(N)$:

t	0	1	2	3	4
N	10	20	40	80	160
$y = \ln(N)$	$\ln(10)$	$\ln(20)$	$\ln(40)$	$\ln(80)$	$\ln(160)$
$y \approx$	2.303	2.996	3.689	4.382	5.075

2. Justification : Les données N suivent une progression géométrique ($N = 10 \times 2^t$), ce qui est exponentiel. Le logarithme transforme la relation exponentielle en relation linéaire : $\ln(N) = \ln(10) + t\ln(2)$.
3. Déterminer la droite $y = at + b$:
   Moyennes :
$$\bar{x} = \frac{0+1+2+3+4}{5} = 2, \quad \bar{y} \approx \frac{2.303 + 2.996 + 3.689 + 4.382 + 5.075}{5} \approx 3.689$$
   Pente :
$$a = \frac{\mathrm{Cov}(t, y)}{V(t)} \approx \frac{1.1}{2} \approx 0.693$$
   Ordonnée à l'origine :
$$b = \bar{y} - a \cdot \bar{x} \approx 3.689 - 0.693 \times 2 \approx 2.303$$
   Droite : $y \approx 0.693t + 2.303$ (note : $\ln(2) \approx 0.693$)
4. Déduction : Puisque $y = \ln(N)$:
$$\ln(N) = 0.693t + 2.303$$
$$N = e^{0.693t+2.303} = e^{2.303} \cdot e^{0.693t} \approx 10 \cdot e^{0.693t} = 10 \cdot 2^t$$

Correction 5 — Problème concret () [Énoncé]

Justification du modèle exponentiel :

1. Motivations économiques : Les dépenses publicitaires peuvent avoir des rendements décroissants. Initialement, chaque euro supplémentaire en publicité augmente beaucoup les ventes, puis l'effet s'atténue (saturation du marché).
2. Modèle exponentiel : $\mathrm{CA} = A \cdot e^{k \cdot \mathrm{Publicit\acute{e}}}$ ou décroissant $\mathrm{CA} = A \cdot (1 - e^{-k \cdot \mathrm{Publicit\acute{e}}})$.
3. Validation du modèle :
   - Collecter les données : dépenses publicitaires et chiffre d'affaires correspondant
   - Appliquer la transformation $z = \ln(\mathrm{CA})$ ou ajuster directement
   - Vérifier la qualité de l'ajustement par :
     - Coefficient de corrélation r proche de 1 (ou R^2 proche de 1 pour ajustement exponentiel)
     - Résidus faibles et sans structure systématique
     - Graphique : nuage proche de la courbe ajustée
   - Comparer avec un modèle linéaire pour voir quel est le meilleur