

**Campus Ynov Aix-en-Provence - M1 IA/Data**

**Module « Data science »**

**Février 2026**

Nom : \_\_\_\_\_ Prénom : \_\_\_\_\_

---

*Cochez les affirmations vraies. Il peut y avoir plusieurs réponses correcte pour une question.  
Pour info une des questions n'a aucune réponse correcte.*

*Cotation : +1 par réponse correcte, malus de -1 par réponse fausse avec un plafond de 25% des réponses. Interdit de cocher toutes les cases d'une question (malus supplémentaire de -1).*

---

**1. Parmi les éléments suivants, lesquels sont considérés comme des données personnelles au sens du RGPD ?**

- A. Une adresse IP
- B. Un numéro de client interne
- C. La température moyenne d'une ville
- D. Une combinaison code postal + date de naissance + sexe

**2. Un Data Scientist utilise un dataset public pour entraîner un modèle de scoring crédit. Le dataset contient l'âge, le code postal et le statut d'emploi. Aucune variable "origine ethnique" n'est présente. Quelle affirmation est correcte ?**

- A. Le modèle ne peut pas être discriminatoire puisqu'aucune donnée sensible n'est utilisée
- B. Ce modèle risque d'induire un biais discriminatoire
- C. L'utilisation d'un dataset public dispense de toute obligation RGPD
- D. Seules les données directement identifiantes posent problème pour la conformité

**3. Soit le noyau suivant :**

$$K = \begin{vmatrix} 1 & 0 & -1 & 0 \\ | & -1 & 5 & -1 \\ | & 0 & -1 & 0 \end{vmatrix}$$

**Quel est l'effet de ce filtre ?**

- A. Flou gaussien
- B. Détection de contours
- C. Accentuation de la netteté (sharpen)
- D. Effet de relief (emboss)

**4. Quelle est la différence entre le top-hat et le black-hat en morphologie mathématique ?**

- A. Le top-hat réalise l'opération (Original – Ouverture), le black-hat réalise l'opération (Fermeture – Original)
- B. Le top-hat utilise l'ouverture, le black-hat utilise la fermeture, mais les deux extraient des détails de même nature
- C. Le top-hat extrait les petits éléments clairs, le black-hat extrait les petits éléments sombres
- D. Ils sont identiques mais appliqués dans des ordres différents

**5. Quelle est la principale différence entre un processus AR(p) et un processus MA(q) ?**

- A. AR utilise les valeurs passées de la série, MA utilise les erreurs passées
- B. AR modélise la tendance, MA modélise la saisonnalité
- C. AR est toujours stationnaire, MA ne l'est jamais
- D. AR a une mémoire infinie, MA a une mémoire finie de q périodes

**6. Vous modélez une série de ventes mensuelles et obtenez les résultats suivants :**

Modèle	AIC	P-value (Ljung-Box)	MAPE (forecast)
ARIMA(1,1,1)	450	0.15	8.5%
SARIMA(0,1,1)(0,1,1)[12]	420	0.42	6.2%
SARIMA(1,1,1)(1,1,1)[12]	415	0.08	6.0%

**Quel modèle choisissez-vous et pourquoi ?**

- A. ARIMA(1,1,1) car il est le plus simple avec la meilleure prédition
- B. SARIMA(0,1,1)(0,1,1)[12] car meilleur AIC et bon test Ljung-Box
- C. SARIMA(1,1,1)(1,1,1)[12] car meilleur AIC
- D. On ne peut pas décider avec ces informations

**7. Lors d'un entretien de cadrage, un responsable marketing vous dit : "On veut de l'IA pour mieux connaître nos clients." Après plusieurs questions, vous découvrez que son vrai problème est que les campagnes emailing ont un taux d'ouverture de 5% et qu'il est sous pression de sa direction. Quelle problématique data science est la plus pertinente ?**

- A. Clustering de clients pour créer des personas marketing
- B. Système de recommandation produits personnalisé
- C. Modèle de scoring pour cibler les clients les plus susceptibles d'ouvrir les emails
- D. Dashboard de connaissance client 360°

**8. Soit les vecteurs  $u = [2, -1, 3]$  et  $v = [1, 4, -2]$ . Quelle est la norme euclidienne de  $u + v$  ?**

- A.  $\sqrt{22}$
- B.  $\sqrt{26}$
- C. 5
- D.  $\sqrt{30}$

**9. L'inertie intra-cluster mesure :**

- A. La distance entre les centroïdes de différents clusters
- B. La somme des distances au carré de chaque point à son centroïde de cluster
- C. Le nombre de points dans chaque cluster
- D. La variance entre les clusters

**10. L'ACP possède les propriétés suivantes :**

- A. L'ACP centre automatiquement les données
- B. Les composantes principales sont orthogonales entre elles
- C. L'ACP peut augmenter le nombre de dimensions
- D. L'ACP est une transformation linéaire

**11. Dans la formule du seuillage d'Otsu, on cherche à minimiser :**

- A. La variance inter-classe
- B. La variance intra-classe
- C. La moyenne des intensités
- D. Le nombre de pixels blancs

**12. Dans l'ACP, la première composante principale est :**

- A. La variable avec la plus grande variance
- B. Le vecteur propre associé à la plus grande valeur propre de la matrice de covariance
- C. La moyenne des variables
- D. Le premier vecteur de la base canonique

**13. Parmi les données suivantes, lesquelles sont considérées comme des données sensibles (catégories particulières) au sens de l'article 9 du RGPD ?**

- A. Le salaire d'un employé       B. L'appartenance syndicale
- C. L'orientation sexuelle       D. Le numéro de sécurité sociale

**14. Soit A une matrice  $3 \times 2$  et B une matrice  $2 \times 4$ . Quelle est la dimension du produit AB ?**

- A.  $3 \times 4$        B.  $2 \times 2$        C.  $3 \times 2$        D. Le produit n'est pas défini

**15. Quelle méthode utilise-t-on couramment pour choisir le nombre optimal de clusters k dans K-means ?**

- A. Méthode du gradient       B. Méthode du coude
- C. Validation croisée       D. Test t de Student

**16. Soit un cluster contenant 3 points : [0, 0], [2, 0], [0, 2]. Quel est le centroïde de ce cluster ?**

- A. [1, 1]       B. [2/3, 2/3]       C. [0, 0]       D. [1, 0]

**17. Quelle affirmation est vraie concernant la relation entre ARIMA et ARMA ?**

- A. ARIMA est toujours plus performant qu'ARMA
- B. ARIMA(p, 0, q) est mathématiquement équivalent à ARMA(p, q)
- C. ARMA ne peut pas modéliser de séries avec tendance, contrairement à ARIMA
- D. ARIMA nécessite toujours au moins une différenciation ( $d \geq 1$ )

**18. Dans l'égalisation d'histogramme, la LUT (Look-Up Table) est calculée à partir de :**

- A. La moyenne des intensités       B. L'histogramme cumulé (CDF)
- C. La dérivée de l'histogramme       D. Le maximum de l'histogramme

**19. Quelle opération morphologique permet de supprimer les petits objets blancs (bruit) tout en préservant les grands objets ?**

- A. Dilatation       B. Érosion       C. Ouverture       D. Fermeture

**20. Une matrice de covariance  $\Sigma$  de dimension  $4 \times 4$  contient :**

- A. 4 variances et 6 covariances uniques       B. 4 variances et 12 covariances
- C. 16 covariances       D. 4 variances et 16 covariances