**Partie A**

ACP – Analyse en Composantes Principales

1. L’ACP est applicable uniquement aux variables qualitatives. Faux. L’ACP s’applique uniquement aux variables quantitatives, nécessite des variables quantitatives continues pour calculer les distances et corrélations.
2. L’ACP permet de réduire la dimension tout en conservant une grande partie de l’inertie. Vrai. L’ACP extrait les axes principaux qui expliquent le maximum de variance (inertie) des données.
3. Les axes de l’ACP sont orthogonaux. Vrai. Les axes principaux sont des vecteurs propres orthogonaux.
4. L’ACP nécessite toujours un centrage et une réduction des données. Faux. Le centrage et la réduction sont nécessaires si les variables n’ont pas la même unité ou échelle. Cela permet de rendre les variables comparables.
5. L’inertie totale correspond à la somme des valeurs propres. Vrai. Chaque valeur propre correspond à la variance expliquée par un axe.
6. En ACP, un individu proche de l’origine a un comportement moyen. Vrai. L’origine représente le profil moyen des individus.
7. Les variables sont projetées dans le cercle des corrélations. Vrai. Le cercle des corrélations permet de visualiser la représentation des variables sur les axes principaux.
8. Les variables fortement corrélées sont représentées par des vecteurs proches. Vrai. Plus l’angle entre deux vecteurs est faible, plus la corrélation est forte.
9. Une valeur propre proche de zéro signifie que la composante correspondante explique beaucoup d'inertie. Faux. Une valeur propre proche de zéro indique que l’axe n’explique presque aucune variance.
10. L’ACP est insensible aux valeurs aberrantes. Faux. L’ACP est sensible aux outliers qui peuvent influencer les axes principaux.

AFC – Analyse Factorielle des Correspondances

1. L’AFC s’applique à un tableau croisé de fréquences. Vrai. L’AFC analyse des tableaux de contingence.
2. L’AFC suppose des données quantitatives. Faux. Elle s’applique à des données qualitatives résumées sous forme de fréquences.
3. L’AFC maximise la variance totale. Faux. Elle maximise l’inertie (variance pondérée par les effectifs).
4. En AFC, les lignes et colonnes sont projetées dans le même plan factoriel. Vrai. Les deux sont représentées conjointement.
5. Les profils ligne et colonne doivent être centrés autour de leur moyenne pour l’analyse. Vrai. Cela permet de comparer les profils aux profils moyens.
6. Une association forte entre deux modalités se traduit par leur proximité dans le plan factoriel. Vrai. Plus deux modalités sont proches, plus elles sont associées.
7. En AFC, les distances entre les modalités sont mesurées avec la distance euclidienne simple. Faux. On utilise la distance du Khi-2.
8. Un point proche de l’origine indique qu’il a un comportement moyen. Vrai.
9. L’AFC est sensible aux effectifs faibles dans les cases du tableau. Vrai. Les faibles effectifs peuvent fausser l’analyse.
10. Les points les plus extrêmes dans le plan sont toujours les plus représentatifs. Faux. Ils peuvent être atypiques ou peu représentatifs.

AFCM – Analyse Factorielle des Correspondances Multiples

1. L’AFCM est utilisée pour l’analyse de plusieurs variables qualitatives. Vrai.
2. L’AFCM suppose un encodage en variables numériques centrées réduites. Faux. Elle utilise un codage binaire (tableau disjonctif complet).
3. L’AFCM est une généralisation de l’ACP. Faux. C’est une généralisation de l’AFC.
4. Le total de l’inertie est égal au nombre de variables actives. Faux. Il dépend du nombre de modalités moins le nombre de variables.
5. L’AFCM est insensible aux modalités rares. Faux. Les modalités rares peuvent influencer l’analyse.
6. L’AFCM est utile pour les enquêtes avec des questions à réponses fermées. Vrai.
7. Les modalités très proches dans le plan ont un profil similaire. Vrai.
8. Les coordonnées des modalités sont pondérées par leur fréquence. Vrai.
9. L’AFCM permet d'étudier la structure des individus à partir des variables continues. Faux. Elle s’applique aux variables qualitatives.
10. Le plan factoriel en AFCM est interprété comme celui de l’ACP. Vrai. L’interprétation graphique est similaire.

CAH – Classification Ascendante Hiérarchique

1. La CAH construit des groupes d’individus à partir d’une matrice de distances. Vrai.
2. Le dendrogramme permet de visualiser l’ordre des fusions. Vrai.
3. En CAH, l’algorithme commence par un seul groupe contenant tous les individus. Faux. Il commence par autant de groupes que d’individus.
4. Le critère de Ward maximise l'inertie intra-classe. Faux. Il minimise l’inertie intra-classe (et maximise l’inertie inter-classe).
5. Le dendrogramme peut être coupé pour définir un nombre de classes. Vrai.
6. La CAH est sensible à l’échelle des variables. Vrai. Il faut souvent standardiser les données.
7. Les groupes obtenus par CAH sont toujours optimaux. Faux. Ils dépendent de la méthode et peuvent ne pas être optimaux.
8. La CAH peut être combinée à l’ACP pour la visualisation. Vrai.
9. La distance de Manhattan est une des distances possibles pour la CAH. Vrai.
10. L’algorithme de CAH est itératif et non récursif. Vrai.

K-Means – Classification des Moyennes

1. K-Means est une méthode de classification non supervisée. Vrai.
2. Le nombre de groupes doit être défini à l’avance. Vrai.
3. K-Means fonctionne mieux avec des variables qualitatives. Faux. Elle est adaptée aux variables quantitatives.
4. L’algorithme minimise la variance intra-groupe. Vrai.
5. K-Means est sensible aux valeurs extrêmes. Vrai.
6. Le résultat de K-Means dépend des points initiaux choisis. Vrai.
7. K-Means peut produire des clusters non sphériques. Faux. Elle tend à former des clusters sphériques.
8. L’ACP peut être utilisée pour visualiser les résultats de K-Means. Vrai.

ICA – Analyse en Composantes Indépendantes

* L’ICA cherche des composantes orthogonales. Faux. Elle cherche des composantes indépendantes, pas nécessairement orthogonales.
* L’ICA est souvent utilisée pour le traitement du signal. Vrai.

Méthodes multidimensionnelles : concepts avancés

1. En ACP, un axe principal est un vecteur propre de la matrice des corrélations. Vrai. Les axes principaux sont les vecteurs propres de la matrice de covariance ou de corrélations.
2. En ACP, les valeurs propres reflètent la corrélation entre variables. Faux. Les valeurs propres représentent la variance expliquée par chaque axe, pas la corrélation entre variables.
3. En AFC, les masses sont les effectifs totaux des lignes ou des colonnes. Vrai. Les masses correspondent aux effectifs marginaux divisés par l’effectif total.
4. L’AFC peut être vue comme une ACP sur un tableau transformé. Vrai. L’AFC est une ACP appliquée à un tableau de contingence transformé en profils.
5. En CAH, deux classes peuvent être fusionnées même si leur distance est grande. Faux. La CAH fusionne toujours les classes les plus proches.
6. En K-Means, chaque individu est affecté à plusieurs clusters avec une probabilité. Faux. K-Means est une classification dure : chaque individu appartient à un seul cluster.
7. La CAH produit des classes imbriquées (hiérarchiques). Vrai. La CAH construit une hiérarchie de classes imbriquées représentée par un dendrogramme.
8. L’ACP permet de détecter des groupes d’individus sans faire de classification. Vrai. L’ACP met en évidence des regroupements visuels sans classification formelle.
9. L’AFCM permet d’inclure des variables quantitatives dans l’analyse. Faux. L’AFCM s’applique aux variables qualitatives uniquement.
10. L’ICA suppose une indépendance des composantes sources. Vrai. L’ICA cherche à extraire des composantes statistiquement indépendantes.

Logiciel R

1. En R, la fonction prcomp() effectue une ACP. Vrai. prcomp() est une fonction standard pour réaliser une ACP en R.
2. En R, la fonction princomp() est recommandée pour les grands jeux de données. Faux. prcomp() est généralement préférée pour les grands jeux de données.
3. La fonction factoextra::fviz\_pca\_ind() permet de visualiser les individus d'une ACP. Vrai. Cette fonction visualise les individus sur les plans factoriels.
4. Le package FactoMineR permet de faire de l’ICA en R. Faux. FactoMineR est dédié aux analyses factorielles, pas à l’ICA.
5. En R, cluster::agnes() permet de faire une CAH. Vrai. agnes() réalise une classification ascendante hiérarchique.
6. En R, l’AFC se fait avec CA() de FactoMineR. Vrai. CA() réalise une Analyse Factorielle des Correspondances.
7. En R, ICASSO est utilisé pour stabiliser les résultats d’ICA. Vrai. ICASSO évalue la stabilité des composantes indépendantes.

**Partie B : exercices sur le choix des méthodes statistiques appropriées**

Applications en biologie/agronomie

1. L’ACP peut être utilisée pour explorer des données morphométriques d’organismes. Vrai. L’ACP est couramment utilisée en morphométrie.
2. L’AFC peut être utilisée pour explorer la relation entre espèces et milieux. Vrai. L’AFC analyse les tableaux de contingence entre espèces et habitats.
3. L’AFCM est adaptée pour étudier les préférences de consommateurs sur des variétés de fruits. Vrai. L’AFCM traite plusieurs variables qualitatives.
4. La CAH permet de classifier des parcelles agricoles selon des variables environnementales. Vrai. La CAH regroupe des unités selon leurs caractéristiques.
5. L’ICA est souvent utilisée pour la séparation de signaux EEG ou EMG. Vrai. L’ICA est largement utilisée en traitement de signal biomédical.
6. K-Means permet de classer des génotypes selon leur performance dans différents environnements. Vrai. K-Means regroupe des individus selon leurs caractéristiques quantitatives.
7. Les analyses factorielles sont peu pertinentes dans les essais agronomiques. Faux. Elles sont très utiles pour réduire la dimension et interpréter les données complexes.
8. En écologie, l’AFC aide à identifier des gradients environnementaux. Vrai. L’AFC met en évidence des associations entre modalités et gradients.
9. En biologie, les axes factoriels sont souvent interprétés comme des gradients de sélection. Vrai. Les axes peuvent représenter des facteurs écologiques ou génétiques.
10. L’ACP est utilisée pour détecter des outliers génétiques. Vrai. Les individus éloignés dans le plan ACP sont souvent des outliers.

**Partie C**

Erreurs fréquentes et précautions

1. Il faut toujours garder le maximum d’axes dans une ACP. Faux. Il faut sélectionner un nombre limité d’axes expliquant une part significative de la variance.
2. L’interprétation graphique repose uniquement sur la proximité entre points. Faux. L’interprétation prend aussi en compte les contributions et cos² des points.
3. Le premier axe en ACP explique toujours plus de 50 % de la variance. Faux. Cela dépend des données, le premier axe peut expliquer moins.
4. En AFC, les modalités rarement présentes doivent être ignorées systématiquement. Faux. Elles peuvent contenir des informations importantes.
5. En K-Means, les clusters sont toujours bien séparés visuellement. Faux. Les clusters peuvent se chevaucher ou être proches.
6. L’ordre des colonnes d’un tableau de données change les résultats d’ACP. Faux. L’ACP est insensible à l’ordre des variables.
7. En AFCM, il est inutile de recoder les modalités absentes. Faux. Les modalités absentes doivent être codées pour éviter des biais.
8. Une modalité fortement corrélée avec un axe est bien représentée sur ce dernier. Vrai. Une forte corrélation signifie une bonne qualité de représentation.
9. L’ACP peut être utilisée avant une CAH pour réduire le bruit. Vrai. L’ACP réduit la dimension et le bruit avant classification.
10. L’AFC est équivalente à une ACP sur un tableau de fréquences. Vrai. L’AFC est une ACP adaptée aux tableaux de contingence.

Statistiques et interprétations

1. En ACP, un cos² élevé signifie une bonne qualité de représentation. Vrai. Le cos² mesure la qualité de projection d’un point sur un axe.
2. L’angle entre vecteurs de variables en ACP reflète leur corrélation. Vrai. Plus l’angle est petit, plus la corrélation est forte.
3. Un individu éloigné de l’origine dans le plan ACP est atypique. Vrai. Il présente un profil différent de la moyenne.
4. En AFC, l’inertie totale est égale au chi² du tableau rapporté à l’effectif total. Vrai. L’inertie est la statistique χ² divisée par l’effectif total.
5. En AFCM, les contributions permettent d’identifier les modalités qui structurent l’axe. Vrai. Les contributions mesurent l’influence des modalités sur les axes.
6. En CAH, une coupe basse du dendrogramme donne peu de classes. Faux. Une coupe basse donne beaucoup de classes (plus fine).
7. Le K optimal en K-Means peut être estimé avec la méthode du coude. Vrai. La méthode du coude aide à choisir le nombre de clusters.
8. L’ICA est robuste aux rotations de l’espace initial. Faux. L’ICA est sensible à la rotation, contrairement à l’ACP.
9. Les axes en AFCM sont orthogonaux. Vrai. Les axes factoriels sont construits orthogonalement.
10. L’AFCM impose que toutes les modalités aient les mêmes poids. Faux. Les modalités sont pondérées selon leur fréquence.