ANALYSE DE DONNÉES MULTIVARIÉE : ACP, AFC, ACM ET CAH

Durée & Format

Durée : 28 heures

Format disponible : Présentiel ou Distanciel

Tarifs

Tarif public intra : Consulter notre site Internet

Tarif public inter : Consulter notre site Internet

Public visé

- Toute personne souhaitant comprendre la structure théorique des méthodes d'analyses factorielles multivariées (ACP, AFC, kmeans, CAH, …) utilisées en Data Mining afin d'en maîtriser la mise en application et l'interprétation des résultats.

Prérequis

- Il est nécessaire que les participants aient de bonnes connaissances sur les outils statistiques de base : corrélation, écart-type, variance, intervalles de confiance, tests d'hypothèses.

Objectifs pédagogiques

A l'issue de cette formation, l'apprenant sera capable de :

- Identifier quelle méthode d'analyse multivariée ou de classification utiliser selon le contexte
- Décrire les concepts mathématiques inhérents à ces méthodes
- Mettre en œuvre les analyses factorielles multivariées suivantes : ACP, AFC, CAH, KMEANS, AFCM, AFD
- Comprendre le rôle de chacun de ces outils et le contexte de leurs utilisations
- Interpréter les résultats et les graphiques qui découlent des analyses ci-dessus
- Interpréter les coefficients et les paramètres permettant d'estimer la qualité des analyses statistiques ci-dessus
- Savoir lire les graphiques issus des analyses multivariées (cercle des corrélations, individus, graphe des éboulis, dendrogrammes, …)
- Comprendre la notion de contributions dans les analyses multivariées
- Comprendre les calculs de distances et les différentes méthodes d'agrégation les plus adaptées à l'objectif de classification
- Comprendre les différences entre la CAH et les k-Means
- Mettre en œuvre et interpréter les résultats d'une méthode de classification de type AFDIdentifier quelle méthode d'analyse multivariée ou de classification utiliser selon le contexte

Programme et déroulement

Jour 1 – Matin

Généralités sur les différentes méthodes d'analyses multidimensionnelles

- Limites des statistiques classiques
- Champs d'application des différentes méthodes d'analyses multidimensionnelles
- Introduction sur le data mining - Les objectifs
  - Objectifs de description

    - Objectifs de prédiction
- Structure des jeux de données
- Présentation de l'éventail des méthodes
    - Analyse en composantes principales
    - Analyse factorielle des correspondances simples et multiples
    - Analyse canonique des corrélations
    - Analyse factorielle discriminante
    - Méthodes de classification : classification ascendante hiérarchique, k-means

Jour 1 – Après-midi

- Principes généraux des différentes méthodes - Notions de :
    - Distance,
    - Inertie et variance
    - Axes factoriels

Jour 2 – Matin

Notion de corrélation

- Définition du coefficient de corrélation
- Interprétation de la valeur du coefficient de corrélation
- Les confusions : corrélation, causalité, pente…
- Les différents coefficients de corrélation
    - Coefficient de Pearson
    - Coefficient de Spearman

Mise en œuvre d'une analyse en composante principales (ACP)

- Structure du jeu de données et contexte d'application
- Objectifs détaillés de l'ACP
- Choix des axes de représentation (choix du nombre de composantes principales)
- Interprétation des sorties graphiques : cercle factoriel et graphique des individus
- Interprétation des axes factoriels
- Contribution des individus et des variables aux axes
- Qualité de représentation des individus et des variables sur les axes : les cosinus carrés

Jour 2 – Après-midi

- Les différentes ACP :
    - ACP normée et non normée
    - ACP non paramétrique
- Positionnement d'une variable illustrative catégorielle
- Les confusions et erreurs à ne pas commettre
- Présentation rapide des Analyses des correspondances simple et multiple si souhaité
- Différences entre ACP et AFC

Jour 3 – Matin

Mise en œuvre d'une analyse factorielle des correspondances (AFC)

- Structure du jeu de données : tableau de contingence, données individuelles (variables qualitatives)
- Contexte d'application et objectifs détaillés de l'AFC
- Différence entre ACP et AFC
- Méthodologie de l'AFC

    - Distance du Chi²
    - Profils lignes
    - Profils colonnes
- Choix des axes de représentation (choix du nombre d'axes)
- Interprétation des sorties graphiques : graphique des modalités
- Contributions et cosinus carrés des modalités aux axes
- Les confusions et erreurs à ne pas commettre

Jour 3 – Après-midi

Mise en œuvre d'une analyse des correspondances multiples (ACM ou AFCM)

- Structure du jeu de données
- Contexte d'application et objectifs détaillés de l'AFCM
- Différence entre AFC et AFCM
- Méthodologie de l'AFCM
- Choix des axes de représentation (choix du nombre d'axes)
- Correction de Benzecri : valeurs propres et taux d'inertie expliquée corrigés
- Interprétation des sorties graphiques :
    - Graphique des modalités
    - Graphique des variables
- Contributions et cosinus carrés des modalités aux axes
- Les confusions et erreurs à ne pas commettre

Jour 4 – Matin

Mise en œuvre d'une classification par une classification ascendante hiérarchique (CAH)

- Structure du jeu de données
- Contexte d'application et objectifs détaillés de la CAH
- Lecture d'un dendrogramme
- Choix du nombre de classes
- Classification sur les individus
- Classification sur les variables
- Classification sur les modalités
- Choix de la distance selon le contexte : classification sur les individus, les variables ou les modalités
- Choix du linkage (saut, critère de « rapprochement » entre les classes) : méthode de Ward, lien simple, lien complet…
- Interprétation des sorties logiciel

Mise en œuvre d'une classification par une k-means

- Présentation des objectifs de la méthode des k-means
- Avantages et inconvénients de la CAH et des k-means
- Détermination des clusters
- Présentation des différentes versions de l'algorithme
- Utilisation de la k-means en complément de l'ACP
- Classification sur grand jeu de données
- Conseils de mise en œuvre
- Interprétation des sorties logiciel

Jour 4 – Après-midi

Mise en œuvre d'une analyse factorielle discriminante (AFD)

- Structure du jeu de données et contexte d'application
- Objectifs détaillés de l'AFD
- Notions de classement et de discrimination
- Méthodologie de l'AFD
- Comparaison avec l'ACP
- Interprétation des sorties logiciel : cercle factoriels, corrélations variables x axes
- Qualité de l'AFD (de la discrimination obtenue)
  - Tests univariés et multivariés (lambda de Wilks)
  - Graphique des individus
  - Matrice de confusion (et éventuellement courbe ROC)
- Les confusions et erreurs à ne pas commettre

JEUX DE DONNEES

Afin de s'approcher au mieux des réalités quotidiennes des praticiens, nous suggérons de nous appuyer pour l'animation pratique de thématiques et surtout de jeux de données reflétant le quotidien des apprenants.

Cet élément est un facteur de réussite pour la formation. Elle permet aux apprenants de :

- Se "reconnaitre" dans les thèmes abordés
- Mieux percevoir l'intérêt des notions étudiées
- S'approprier le contenu de la formation

Il sera donc pertinent que les apprenants puissent réfléchir en amont de la formation à des problématiques, jeux de données ou documents susceptibles d'être utilisés en support lors de la formation.

OUTIL LOGICIEL

Cette formation n'est pas strictement dédiée à un logiciel. Les exercices et les illustrations se feront à l'aide d'un logiciel "spécialisé" pour ce type d'analyses : R, Jamovi, ...

On notera la possibilité d'utiliser Jasp, logiciel open source libre et gratuit, très utilisé dans la communauté scientifique.

Le logiciel retenu sera choisi en accord avec les participants.

Sur le logiciel R, nous proposons de nous appuyer particulièrement sur le package factominer avec ses satellites (factoshiny, ...).

Certification

Consulter le programme sur notre site internet pour identifier si cette formation est certifiante.

Modalités pédagogiques

- Explications théoriques suivies de pratiques guidées puis de mises en autonomie.
- Exercices autonomes et réguliers pour assurer l'assimilation

Moyens et supports pédagogiques

Méthodes et moyens

- Evaluation préformation pour définir ses besoins et ses priorités d'usage de l'outil
- Approche participative des connaissances à transmettre
- Mise en situation concrète des outils et méthodes
- Adaptation du rythme et des exercices au niveau du groupe
- Utilisation des 3 vecteurs de communication : oral (vulgarisation du vocabulaire), visuel (vidéoprotection du logiciel) et kinesthésique (travaux pratiques)

Support stagiaire

- Support papier ou électronique (dématérialisé)
- Les exercices d'accompagnement peuvent être récupérés sur clef USB
- Exercices supplémentaires téléchargeables sur notre site

Evaluation et suivi

Les objectifs pédagogiques sont évalués et suivis grâce à différentes **méthodes adaptées aux acquisitions de compétences visées**. Nos formateurs réalisent ces évaluations **tout au long de la formation**, que ce soit pendant les séances synchrones ou asynchrones. Voici une liste non exhaustive des méthodes d'évaluation pouvant être utilisées en formation :

- **Questionnaire** de vérification de connaissances (Quiz)
- Réalisation de **Travaux Pratiques** : production ou amélioration d'un fichier
- **Mises en situation** et grilles d'analyse
- **Serious Game** : jeu de rôles et analyse
- **Activités interactives** à travers l'utilisation d'une plate-forme connectée
- …

Toutes nos formations intègrent **une auto-évaluation** via notre Extranet Stagiaire au début et à la fin de chaque action de formation. Cet outil offre à chacun la possibilité de mesurer sa progression par rapport aux objectifs pédagogiques visés et leurs atteintes.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire via son extranet stagiaire.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite. En savoir plus