

ANALYSES MULTIVARIÉES, CLASSIFICATION (ACP, AFC, ACM, CAH, K-MEANS, AFD)

Cette formation s'adresse à des personnes souhaitant maîtriser les concepts et la mise en œuvre des analyses factorielles multivariées. Ces analyses ont pour objectif d'extraire des informations de données :

- *Volumineuses en nombre de variables*
- *Volumineuses en nombre d'individus*
- *Non structurées*
- *Comportant des variables redondantes (confusions entre variables)*

La formation conviendra tout à fait à un public venant chercher du savoir statistique sur :

- *Les concepts des différentes méthodes descriptives d'analyses factorielles multivariées et de classification*
- *Les contextes d'application de chaque méthode*
- *Les fondements mathématiques (méthodologiques) de ces analyses*
- *La mise en œuvre et l'interprétation des résultats des différentes méthodes*

Durée : 28 heures (4 jours)

Lieu : Formation disponible en présentiel dans vos locaux, ou en ligne en visioconférence

Tarifs :

- Formation en ligne inter-entreprise : n/a
- Formation en ligne intra-entreprise : 6 960,00€ HT / formation
- Formation sur site intra-entreprise : 6 960,00€ HT / formation + frais de déplacement du formateur

Profils des stagiaires :

Technicien, employé, cadre, chercheur, étudiant

Prérequis :

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

Objectifs pédagogiques

- 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, AFM, AFD
- Interpréter les résultats et les graphiques qui découlent des analyses ci-dessus
- Maîtriser les coefficients et les paramètres permettant d'estimer la qualité des analyses statistiques ci-dessus
- Expliquer la différence entre les notions de contribution et de cosinus carré
- Expliquer les distances et les 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 CAH et k-means

Contenu de la formation

- **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 - 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
- **Principes généraux des différentes méthodes - Notions de : Distance - Inertie et variance - Axes factoriels**
- **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
 - 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
- **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
- **Mise en œuvre d'une Analyse des Correspondances Multiples (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 expliquées corrigées
 - 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
- **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 du logiciel
- **Mise en œuvre d'une classification par la méthode 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 du logiciel
- **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 du 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

Organisation de la formation

Equipe pédagogique Arkesys :

- Responsable de la filière statistique du groupe Arkesys et formateur en statistique, **Thierry Anthouard** est passionné depuis toujours par le domaine de la statistique. C'est en 1992 qu'il a lancé le développement de la filière formation statistique du Groupe Arkesys. Son approche pédagogique "par l'exemple" lui permet de vulgariser la statistique et de la rendre accessible à tous les apprenants. Consultant dédié à l'accompagnement de clients grands comptes, il s'adapte à tout type de contexte et de problématique d'apprentissage.
- Formateur en statistique, **Jérôme-Philippe Garsi** bénéficie d'une expérience de 13 ans dans le domaine de la formation. Depuis son doctorat sur des questions cliniques, ses travaux visent principalement l'intérêt des populations, leur santé et leur bien-être. A l'aise quel que soit le public, il fait de la pédagogie et de la vulgarisation des connaissances scientifiques un impératif. Pour cela, il fait toujours preuve du plus grand soin de clarté aussi bien dans ses documents écrits que dans ses présentations orales.

Moyens pédagogiques et techniques :

- Accueil des apprenants dans une salle dédiée à la formation.
- Documents supports de formation projetés.
- Les exercices d'accompagnement peuvent être récupérés sur clef USB

Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation :

- Feuilles de présence
- Exercices de synthèse et d'évaluation
- Explications théoriques suivies de pratiques guidées puis mises en autonomie
- Evaluation de fin de stage
- Certificat de réalisation de l'action de formation

Accessibilité aux personnes handicapées :

Les personnes en situation de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

Délais d'accessibilité à la formation avant le démarrage : 2 semaines

Addinsoft

40 rue Damrémont
75018 Paris
Email: training@xlstat.com
Tel: 0170060488



Formation en ligne : Le lien de connexion à la classe virtuelle sera envoyé par email la semaine qui précède la formation.

Contact : Pour toute demande d'informations complémentaires vous pouvez nous contacter par email training@xlstat.com ou par téléphone au 0170060488.