

Inter - Mettre en œuvre des analyses factorielles multivariées (ACP, CAH, K-MEANS, AFD, PLS)

Les 9, 10, 11 et 12 octobre 2023

Objectifs	<ul style="list-style-type: none">• 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 suivantes : ACP, CAH, K-MEANS, AFD, PLS• Maitriser 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• Identifier les contextes d'utilisation des différentes versions de la régression PLS• Comprendre les concepts mathématiques inhérents à la régression PLS• Interpréter les résultats et les graphiques qui découlent des analyses
Public	Ingénieurs, techniciens, chercheurs, doctorants.
Pré-requis	<p>Bonnes connaissances sur les outils statistiques de base : corrélation, écart-type, variance, intervalles de confiance, tests d'hypothèses, régression linéaire.</p> <p>Après votre inscription dans Sirène, compléter le questionnaire suivant : https://forms.office.com/Pages/ResponsePage.aspx?id=SdMr0PDW80WLuFdUfKcV/NhSI5D6av0tFtLyEn6XmsGhUQzZGWkJLMlk1Szg4TUcxUFBQQUVQNIJGTy4u</p> <p><i>Si le lien ne fonctionne pas en cliquant dessus ->copier/coller dans votre barre de recherche Internet</i></p> <p>Sans ce questionnaire complété et votre inscription dans Sirène validé par votre supérieur hiérarchique avant la date limite d'inscription, votre demande de formation ne sera pas étudiée</p>
Programme	<p>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 - Structure des jeux de données - Présentation de l'éventail des méthodes - Principes généraux des différentes méthodes</p> <p>Notions 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</p> <p>Structure du jeu de données et contexte d'application</p> <ul style="list-style-type: none">• 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• 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 <p>Mise en œuvre d'une classification par une classification ascendante hiérarchique (CAH)</p> <ul style="list-style-type: none">• Présentation des objectifs de la classification ascendante hiérarchique• Structure du jeu de données• Choix de la distance• Choix du linkage (saut, critère de « rapprochement » entre les classes) : méthode de Ward..• Lecture d'un dendrogramme

- Choix du nombre de classes
- Classification sur les individus
- Classification sur les variables
- 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

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)
- Les confusions et erreurs à ne pas commettre
- Présentation de la PCR (Principal Component Regression)
- Méthodologie générale de PLS
- Présentation des différentes versions de la régression PLS
- Structure du jeu de données et contexte d'application de la PLS-DA
- Objectifs détaillés de la PLS-DA
- Choix du nombre de composante de la PLS : critère du Q carré
- Interprétation des sorties logiciel : cercle factoriels, corrélations variables x axes
- Importance des variables pour la discrimination des groupes : les VIP
- Qualité de la PLS-DA
- Les confusions et erreurs à ne pas commettre

Mise en œuvre d'une régression PLS

- Contexte et objectifs
- Présentation des différentes méthodes de régression sur des données colinéaires : PCR, régression de Ridge et PLS
- Principes et concepts mathématiques inhérents à la PCR et la régression PLS
- Présentation des différentes versions de la régression PLS
- Mise en œuvre et interprétation des résultats : graphiques, coefficients du modèle...
- Choix du nombre de composantes (validation croisée)
- Composantes et coefficients de régression
- Qualité d'ajustement, qualité de prédiction
- Coefficients Q^2 et R^2
- Importance des variables explicatives pour la prédiction
- Sélection de variables

Dates

Date : **Les 9, 10, 11 et 12 octobre 2023**
Inscriptions avant **le 31 août 2023** sur <https://www.sirene.inserm.fr/>

Lieu

DR Paris IdF Sud – Paris 13^e

Contact

Catherine Marcilhac
INSERM – DR Paris IdF Sud
@ : formation.paris11@inserm.fr
☎ : 01.85.55.38.44

**INSERM DR Paris IDF Sud
Formation Continue**
48-50 rue Albert
75013 Paris