

Inter – Analyses multidimensionnelles : analyses factorielle multivariées, classification, PLS

Du 2 au 5 décembre 2019

<p>Objectifs</p>	<ul style="list-style-type: none"> - D'identifier quelle méthode d'analyse multivariée ou de classification utilisés selon le contexte - De connaître les concepts mathématiques inhérents à ces méthodes - De mettre en œuvre les différentes analyses - D'interpréter les résultats de chaque méthode et d'analyser les différents graphiques qui en découlent - De connaître les coefficients permettant d'estimer la qualité de l'analyse statistique - De mesurer la contribution de chaque variable et de chaque individu sur les axes factoriels - D'identifier les contextes d'utilisation des différentes versions de la régression PLS - De comprendre les concepts mathématiques inhérents à la régression PLS et PLS-DA
<p>Public</p>	<p>Toute personne souhaitant comprendre et mettre en œuvre les procédures d'analyses multidimensionnelles aussi bien lorsque l'on a un grand nombre ou un petit nombre d'individus</p>
<p>Pré-requis</p>	<ul style="list-style-type: none"> - IL EST INDISPENSABLE que les participants aient de bonnes connaissances sur les outils statistiques de base : tests d'hypothèses, p-value, risque alpha, régression linéaire. - Joindre le questionnaire complémentaire à votre demande d'inscription
<p>Programme</p>	<p>Connaitre les généralités sur les méthodes d'analyses factorielles et sur la classification</p> <ul style="list-style-type: none"> • Limites des statistiques classiques • Structure des jeux de données • Champs d'application des méthodes d'analyses multifactorielles • Les objectifs des techniques d'analyse <ul style="list-style-type: none"> ✓ Objectifs de description ✓ Objectifs de prédiction • Présentation de l'éventail des méthodes <ul style="list-style-type: none"> ✓ Analyse en composantes principales ✓ Positionnement multidimensionnel ✓ Analyse factorielle des correspondances ✓ Analyse factorielle des correspondances multiples ✓ Analyse canonique des corrélations ✓ Analyse factorielle discriminante ✓ Méthodes de classification : classification ascendante hiérarchique, réallocation dynamique ✓ Régression PLS, PLS2, PLS-DA • Les outils mathématiques <ul style="list-style-type: none"> ✓ Notion de distances ✓ Notion de corrélations <p>Méthodes factorielles multidimensionnelles</p> <p>Mettre en œuvre l'ACP</p> <ul style="list-style-type: none"> • Structure du jeu de données et contexte d'application • Matrice de corrélations, de variance-covariance • Inertie et variance • Choix des axes de représentation (choix du nombre de composantes principales)

- Interprétation des axes factoriels
- Contribution des individus et des variables aux axes
- Qualité de représentation des individus et des variables sur les axes
- Les différentes ACP : normée et non normée, non paramétrique, polychorique
- Représentations graphiques diverses
- Positionnement d'une variable illustrative catégorielle et ellipses de confiance
- Interprétation des sorties logicielles

Mettre en œuvre l'AFC

- Contexte d'utilisation d'une analyse factorielle des correspondances
- Structure du jeu de données
 - ✓ Tableau de contingence
 - ✓ Données individus, variables qualitatives
- Différence entre ACP et AFC
- Proximités et oppositions entre les modalités, liaison entre deux variables qualitatives
- Profils lignes et profils colonnes
- Choix des axes de représentation
- Représentation des modalités de base autour des composantes
- Contributions des modalités à l'inertie
- Interprétation des sorties logicielles

Mettre en œuvre l'ACC

- Contexte d'utilisation d'une analyse canonique des corrélations
- Structure du jeu de données
- Analogie avec l'ACP, la régression linéaire et les autres méthodes d'analyse factorielle
- Objectifs et principes de l'ACC
- Notion de proximité entre 2 groupes de variables quantitatives
- Vocabulaire spécifique : variables canoniques, coefficients de corrélation canonique
- Représentation des variables et des individus dans les sous-espaces de chaque groupe

Méthodes de classification, de classement

Mettre en œuvre l'AFD

- Contexte d'utilisation de l'analyse factorielle discriminante
- Structure du jeu de données
- Notions de classement et de discrimination
- Méthodologie de l'AFD
- Comparaison avec l'ACP
- Interprétation des sorties logicielles

Mettre en œuvre la CAH

- Présentation des objectifs de la classification ascendante hiérarchique
- Structure du jeu de données
- Choix de la distance, dissimilarité, similarité entre les individus : distance euclidienne...
- 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 (distance de corrélation et composantes latentes)
- Interprétation des sorties logicielles

Mettre en œuvre la classification par réallocation dynamique (centres mobiles)

- Contexte d'utilisation de la classification par réallocation dynamique
- Principe général des différentes méthodes

	<ul style="list-style-type: none"> • Comparaison avec la CAH • Présentation de la méthode k-means • Les différentes variantes de classification par réallocation dynamique • Interprétation des sorties logicielles • Conseils de mise en œuvre <p><u>Régression PLS</u></p> <p>Notions fondamentales de la régression PLS</p> <ul style="list-style-type: none"> • Contexte d'utilisation • Présentation des différentes méthodes de régression sur des données colinéaires : PCR, régression de Ridge et PLS • Notion de variables latentes • Principes et concepts méthodologiques inhérents à la PCR et la régression PLS • Présentation des différentes versions de la régression PLS <p>Mettre en œuvre la régression PLS</p> <ul style="list-style-type: none"> • Analyse des sorties logicielles : graphiques, coefficients du modèle... • Composantes et coefficients de régression • Qualité d'ajustement, qualité de prédiction (validation croisée) • Coefficients Q^2 et R^2 • Importance des variables explicatives pour la prédiction : VIP et sélection des variables explicatives • Choix du nombre de composantes • Mise en œuvre et analyse des résultats des différentes méthodes 		
Dates et lieu	Date : Du 2 au 5 décembre 2019 Lieu : Paris 13 Inscription avant le : 8 octobre 2019 sur https://www.sirene.inserm.fr/		
Contact	<table> <tr> <td data-bbox="472 1211 810 1301"> Nathalie SUZANNE Responsable Formation nathalie.suzanne@inserm.fr </td> <td data-bbox="1027 1211 1262 1301"> Inserm DR Paris 11 48-50 rue Albert 75013 PARIS </td> </tr> </table>	Nathalie SUZANNE Responsable Formation nathalie.suzanne@inserm.fr	Inserm DR Paris 11 48-50 rue Albert 75013 PARIS
Nathalie SUZANNE Responsable Formation nathalie.suzanne@inserm.fr	Inserm DR Paris 11 48-50 rue Albert 75013 PARIS		