



SÉMINAIRE INED DE STATISTIQUE APPLIQUÉEⁱ

Analyse de Données : Nouvelles méthodes, nouveaux outils Séance du jeudi 5 juin 2008 de 14h - 17h 30

Pour cette troisième session, nous nous intéresserons au développement et à l'implémentation de méthodes récentes issues de l'analyse des données. Les méthodes multi-tableaux seront au centre de ces exposés, avec notamment une présentation de l'Analyse Factorielle Multiple, développée initialement par B. Escofier et J. Pagès et l'Analyse en Composantes Thématiques, extension des méthodes d'analyse de données et méthodes PLS.

La présentation du module FactoMineR permettra de montrer un outil convivial axé sur les méthodes d'analyse de données, développé pour le logiciel R, « boîte à outils statistique » issue du logiciel libre et véritable alternative aux logiciels statistiques usuels, notamment pour les milieux de l'Enseignement et de la Recherche.

RÉSUMÉS DES INTERVENTIONS

Elisabeth MORAND, INED

Une introduction à l'Analyse Factorielle Multiple.

L'analyse Factorielle Multiple (AFM) (Escofier et Pagès, 1982) est une méthode permettant d'étudier des individus caractérisés par des groupes de variables de même nature ou de nature différente.

La comparaison de configurations (ou nuages) de points homologues est un problème classique en analyse de données. Historiquement, la première étape a consisté en la comparaison de deux configurations par un ajustement sans déformation au moyen de translations, de transformations orthogonales (Green, 1952, Cliff, 1966) et d'homothéties (Schönemann et Carroll, 1970). Pour comparer un ensemble de plus de deux configurations de points homologues, Kristof et Wingersky (1971) ont proposé un algorithme permettant d'ajuster les configurations entre elles par des translations et des transformations orthogonales. L'Analyse Procrustéenne Généralisée (Gower, 1975), STATIS (Lavit, 1988) ou l'Analyse Factorielle Multiple (Escofier et Pagès, 1982) sont aussi des outils de comparaison de configurations de points homologues.

On s'attachera ici à présenter la structure de tableaux multiples et les principes de l'Analyse Factorielle Multiple. Dans un deuxième temps, on proposera un exemple d'étude de trajectoire à l'aide de l'AFM.

François HUSSON, IRMAR Rennes (Agrocampus – Laboratoire Mathématiques Appliquées).

FactoMinerR : Un package R d'analyse exploratoire des données pour l'Enseignement et la Recherche

R est un logiciel de statistique gratuit en plein essor car utilisé de plus en plus, non seulement par les universitaires mais également par les professionnels. De nombreuses bibliothèques ou packages sont également disponibles permettant d'augmenter les capacités de R. Dans cet exposé, nous montrerons les grands principes du package FactoMineR, package qui est dédié à l'analyse de données à la française. Ainsi, il est possible de faire des analyses factorielles simples (ACP, AFC, ACM) ou d'utiliser des méthodes plus avancées Analyse Factorielle Multiple (AFM), AFM Hiérarchique, AFM Duale). Ce package permet de construire de nombreux graphes, d'ajouter des éléments illustratifs dans les analyses (variables quantitatives et/ou qualitatives supplémentaires, individus supplémentaires), etc. Nous présenterons également l'interface graphique qui est également disponible et qui permet d'utiliser ce package de façon conviviale.

Traitement d'un exemple d'Analyse Factorielle Multiple avec FactoMinerR.

Dans cet exposé, nous montrerons à travers quelques exemples l'intérêt de prendre en compte une structure sur les variables grâce à l'Analyse Factorielle Multiple (AFM). Un exemple sera traité en détail avec le package FactoMineR et nous commenterons les différentes sorties du package.

Xavier BRY, UNIVERSITÉ MONTPELLIER 2 : •

Modélisation explicative multi-tableaux : nouvelles méthodes géométriques.

Les multi-tableaux (juxtapositions de groupes de variables décrivant un même ensemble d'observations) sont couramment traités par deux types de méthodes d'analyse multidimensionnelle fondées sur l'optimisation d'un critère global : les méthodes factorielles, purement exploratoires (STATIS, AFM, ACP, ACPG...) qui extraient plusieurs dimensions par groupe mais ne peuvent les articuler dans un modèle explicatif et les méthodes explicatives à base d'équations structurelles (LISREL), qui n'extraient qu'une dimension par groupe et ne permettent donc d'explorer ceux-ci.

La méthodologie PLS tente un compromis entre les deux approches afin d'en pallier les carences, mais ne sait traiter via un critère global le cas général de plusieurs groupes dont de multiples dimensions sont reliées par un modèle explicatif. Nous proposons deux extensions de la régression PLS permettant de faire face à cette situation. La première (Structural Equation Exploratory Regression) est fondée sur la maximisation d'une covariance multiple. La seconde (Analyse en Composantes Thématiques) ne l'est pas, mais est plus simple à coupler à un modèle linéaire généralisé. La possibilité d'utilisation de ces méthodes en analyse biographique sera décrite et discutée.

Arnaud Bringé et Bénédicte Garnier

ⁱ Institut national d'études démographiques • Établissement public scientifique et technologique

133, bd Davout, Paris 20^e • Standard : 01 56 06 20 00 • Fax : 01 56 06 21 99

Le service Méthodes Statistiques de l'Ined organise un séminaire trimestriel, se proposant de couvrir un domaine de Statistique Appliquée traitant de l'application d'une méthodologie.

Ce séminaire est ouvert à tous, sans frais de participation. Pour tout autre renseignement contacter Madame Chafika Mekhazni, secrétaire du service, au 01 56 06 20 91.