

Cours d'Analyse des données

A. Fischer

Descriptif : 2h de cours + 3h de TP par semaine

Les données sont des résultats d'expériences ou d'enquêtes mesurés, observés, sur un certain nombre d'individus. Il s'agit soit de nombres (variable quantitative), soit de codes (variable qualitative). L'analyse des données est un outil perfectionné de statistique descriptive, qui consiste à étudier un jeu de données individus x variables en recherchant notamment s'il existe des relations entre individus et entre variables.

On peut distinguer 3 groupes de méthodes : la statistique descriptive classique qui permet l'étude d'une ou deux variables observées sur un ensemble d'individus, des analyses portant sur des nuages de points de plus grande dimension, ainsi que la classification automatique consistant à regrouper des individus en catégories homogènes relativement à tel ou tel critère.

Comme les tableaux de données peuvent être très grands en pratique, ce qui nécessite des calculs sur ordinateurs, le module comprend des séances de Travaux Pratiques avec le logiciel R.

Plan du cours :

- 1) Statistique descriptive univariée : variable qualitative, variable quantitative, tableaux des effectifs et des fréquences, représentations graphiques, indicateurs statistiques, test de gaussianité.
- 2) Statistique descriptive bivariée : étude d'une liaison pour un couple de variables quantitative-quantitative, quantitative-qualitative, qualitative-qualitative
- 3) Analyse en composantes principales : résolution mathématique et interprétation
- 4) Analyse factorielle des correspondances : analyse de correspondances simples, analyse des correspondances multiples, Introduction à l'analyse factorielle multiple
- 5) Classification hiérarchique : méthode ascendante
- 6) Analyse factorielle discriminante