

MASTER 1 SCIENCES DU MANGEMENT

Titre: Analyse de données I
Enseignant: Xavier Pieri
Nbre d'heures : 18CM+12TD

Objectifs:

Ce cours s'inscrit dans le domaine de la statistique descriptive : l'objectif est de permettre aux étudiants en économie-gestion de choisir les méthodes adaptées au traitement statistique de leurs données. Les chapitres II et III présentent les principes théoriques fondamentaux pour explorer de vastes tableaux de données quantitatives et qualitatives. Deux méthodes sont abordées dans le programme : l'Analyse en Composantes Principales (ACP) et l'Analyse Factorielles des Correspondances simples (AFC).

Le chapitre IV présente des méthodes et des outils de classifications automatiques qui complètent le panorama de l'analyse descriptive permettant de créer des groupes homogènes d'individus sous forme de typologie de population.

Des données, issues d'enquêtes socio-économiques, illustreront le cours. Les analyses et cas pratiques sont présentés et développés à travers le logiciel SPAD sous windows.

Plan du Cours :

1. Rappels de statistique descriptive élémentaire

1.1. Rappels sur les indicateurs en statistique descriptive. Moment d'une variable, moyenne, écart-type, coefficient de corrélation linéaire, droite d'ajustement linéaire, matrice de covariance, matrice de corrélation de p variables.

2. L'analyse en composante principale (ACP)

2.1. But et objectif de l'ACP.

2.2. Définition d'une distance euclidienne.

- 2.2.1. Distance euclidienne canonique.
- 2.2.2. Distance euclidienne normée.
- 2.2.3. Inertie d'un nuage de points, Inertie due à une droite.

2.3. Théorie de l'analyse en composante principale.

- 2.3.1. Recherche de la droite d'inertie minimum.
- 2.3.2. Recherche du plan d'inertie minimum.
- 2.3.3. Définition des composantes principales.

2.4. Interprétation des résultats d'une ACP.

- 2.4.0. Introduction.
- 2.4.1. Analyse d'une matrice de covariance-corrélation.
- 2.4.2. Etude des valeurs propres.
- 2.4.3. Etude des vecteurs propres.
- 2.4.4. Représentation graphique des variables, cercle de corrélation.
- 2.4.5. Représentation des individus, traitement des individus supplémentaires.

2.5. Initiation à l'utilisation du progiciel statistique SPAD.

- 2.5.1. Introduction de données, description des principaux indicateurs statistiques.

2.5.2. Travaux pratiques: des études de cas seront proposées aux étudiants pour se familiariser au logiciel et à l'ACP.

3. L'analyse factorielle des correspondances (AFC)

3.0. But de l'AFC.

3.1. Principe de la méthode de l'AFC.

- 3.1.1. Notion du profil (ligne et colonne).
- 3.1.2. Notation- distance entre profils.
- 3.1.3. Analyse en composantes principales sur les profils.
- 3.1.4. Analyse duale.

3.2. Comment analyser les résultats d'une AFC.

- 3.2.1. Valeurs propres, pourcentage d'inertie.
- 3.2.2. Représentations graphiques.
- 3.2.3. Contributions absolues et relatives.
- 3.2.4. Représentation simultanée.
- 3.2.4. Travaux pratiques: études de cas en AFC.

4. Méthodes de Classification

- 4.1. Agrégation autour des centres mobiles.
- 4.2. Classification Hiérarchique: Principe. Classification ascendante selon le saut minimal, Critère de Ward généralisé.
- 4.3. Elaboration de typologie d'individus, Choix du nombre de classes. Descriptions statistiques des classes.

Modalités de contrôle : Examen terminal écrit (Cours) et contrôle continu (TD)
