



UNIVERSITE DE NOUAKCHOTT
INSTITUT UNIVERSITAIRE PROFESSIONNEL – IUP
FEPARTEMENT INFORMATIQUE

L3 MAEF



CM1: Analyse des données

Introduction à l'ADD

Dr. EL BENANY Mohamed Mahmoud

Année universitaire : 2024/2025

Objectifs du Module: Analyse des Données

Ce module vise à fournir aux étudiants les fondements théoriques et pratiques des méthodes factorielles en analyse des données, en les initiant aux techniques de réduction de dimension et d'exploration de données multidimensionnelles.

Objectifs du Module

1. Comprendre les concepts clés des méthodes factorielles.
2. Maîtriser la Décomposition en Valeurs Singulières (SVD).
3. Appliquer l'ACP pour réduire et visualiser des données quantitatives.
4. Utiliser l'AFC pour analyser des tableaux de contingence.
5. Conduire une AFCM sur des données catégorielles multiples.
6. Implémenter ces méthodes avec des logiciels (R, Python, SPSS).

Compétences Visées :

1. Exploration des données,
2. réduction de dimension des données
3. interprétation des données.

Concepts fondamentaux de l'analyse des données

Introduction

Chapitre 1 : SVD

Chapitre 2 : Analyse en Composantes Principales

Chapitre 3 : Analyse Factorielle des Correspondances

Chapitre 5 : AFC Multiples (AFCM)

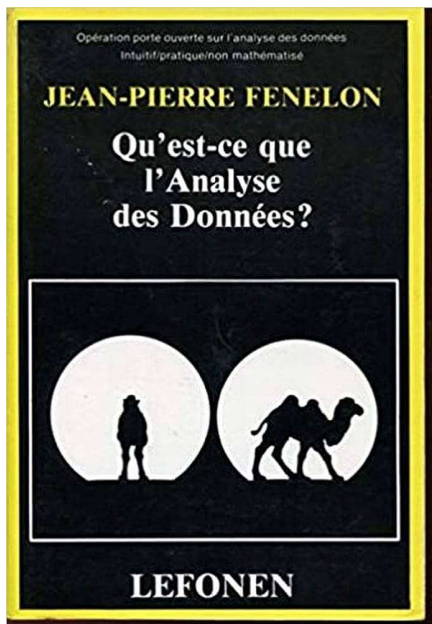
Chapitre 6 : Classification Ascendante Hiérarchique

Conclusion

Introduction

1-Définition et objectifs

L'Analyse des données (ADD) : l'ensemble de méthodes **descriptives** ayant pour objectif de **résumer** et **visualiser** **l'information** contenue dans un grand tableau de données



Selon J-P. Fénelon: « l'analyse des données est un ensemble de techniques pour **découvrir** la structure, éventuellement compliquée, d'un tableau de nombres à plusieurs dimensions et de traduire par une structure plus simple et qui la **résume** au mieux. Cette structure peut le plus souvent, être représentée graphiquement' »

Introduction

1-Définition et objectifs: De la statistique traditionnelle à l'ère du Big Data

Le développement des outils informatiques a fortement contribué au développement de nombreuses méthodes statistiques

Ce qui permet de traiter sans difficultés de vastes données des enquêtes et des investigations (grands tableaux de milliers de lignes et milliers de colonnes)

Outils Modernes :

- Traditionnels (SPSS, EViews)
- Open Source (Python/R)
- Big Data (Spark, Databricks)
- AutoML (H2O.ai, DataRobot)

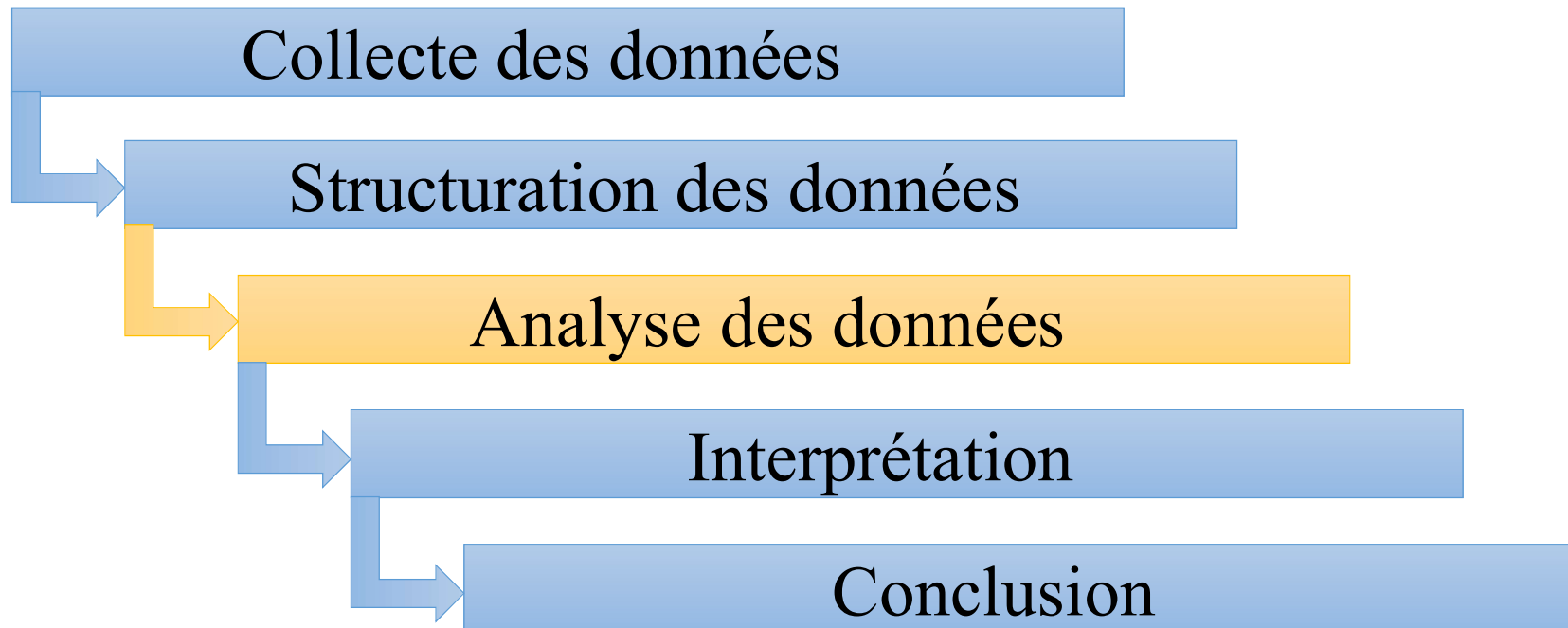
1990-2000 : SPSS, EViews • 10k lignes

2010 : Hadoop • Millions de données

2025 : IA/Cloud • Données illimitées

2- Le processus de ADD

- Les principales étapes du processus d'analyse :



2- Le processus de ADD

Collecte des données

On distingue deux types des données :

Les données primaires sont spécialement collectées pour répondre à une étude statistique précise.

Les données secondaires: sont des données qui existent déjà (Ex: statistiques officielles...)

3- Les méthodes

On distingue plusieurs méthodes selon deux critères

1- Selon l'objectif de la méthode

Méthodes explicatives

Méthodes descriptives

2- Selon type de mesure

**Mesure
nominale**

**Mesure
ordinaire**

**Mesure
métrique**

3- Les méthodes

Méthodes
explicatives

Consiste à expliquer une variable au moyen d'une ou plusieurs variable: (Il s'agit d'une relation de cause à effet en statistiques):

- Variable à expliquer (Y) :** Phénomène que l'on cherche à comprendre/prédire (effet).
- Variables explicatives ($X_1, X_2...$) :** Facteurs qui influencent Y (causes).

Exemple Économétrique:

Équation : $INVES = \alpha \cdot CROI + \beta \cdot INF + \gamma \cdot INDH$

- **INVES (Y) :** Investissements (à expliquer)
- **CROI (X_1) :** Croissance économique (explicative)
- **INF (X_2) :** Inflation (explicative)
- **INDH (X_3) :** Indice de développement humain (explicative)

Interprétation :



Les coefficients α, β, γ mesurent l'impact de chaque facteur sur l'investissement.

Exemple : Si $\alpha = 0.5$, une hausse de 1% de CROI entraîne +0.5% d'INVES.

3- Les méthodes

Méthodes explicatives

Consiste à expliquer une variable au moyen d'une ou plusieurs variable: (Il s'agit d'une **relation de cause à effet** en statistiques):

-  **Variable à expliquer (Y)** : Phénomène que l'on cherche à comprendre/prédire (effet).
-  **Variables explicatives ($X_1, X_2...$)** : Facteurs qui influencent Y (causes).

Exemple :

- **Problèmes de régression et de corrélation**
- **Analyse de la variance**
- **Analyse discriminatoire**
- **Régression logistique**
- **Corrélation canonique**

3- Les méthodes

Consiste à résumer, visualiser et synthétiser les informations.

Méthodes
descriptives

- Exemple :
- Analyse Factorielle des Correspondances
 - Analyse en Composantes Principales
 - Classification Ascendante Hiérarchique

Méthode	Usage	Exemple	Outil
ACP (Analyse en Composantes Principales)	Simplifier des données multidimensionnelles	Réduire 10 variables économiques à 2 axes	Python (sklearn)
AFC (Analyse Factorielle des Correspondances)	Analyser des tableaux de contingence	Lier catégories de produits et régions	R (FactoMineR)
CAH (Classification Ascendante Hiérarchique)	Grouper des individus similaires	Segmenter des clients par comportement	Orange, SPSS
Statistiques descriptives	Résumer une variable	Moyenne, médiane, histogramme	Excel, Tableau

Comparaison vs. Méthodes Explicatives

Aspect	Méthodes Descriptives	Méthodes Explicatives
But	Explorer, visualiser, résumer les données	Prédire ou expliquer des relations de causalité
Question clé	"Comment les données sont-elles structurées ?"	"Pourquoi Y change-t-il quand X varie ?"
Outils typiques	ACP, AFC, CAH, histogrammes, boxplots	Régression linéaire, ANOVA, modèle logistique
Exemple	Segmenter des clients par comportement (CAH)	Prédire les ventes avec le budget pub (régression)
Complexité	Pas d'hypothèses statistiques	Nécessite des hypothèses (ex : linéarité)

3- Les méthodes

Mesure
nominale

On utilise des chiffres sans aucune relation d'ordre, ni de distance, ni d'origine:

- **Femme (1)**
- **Homme (2)**

Exemple :

- **Sexe: Femme , Homme**
- **Situation matrimoniale: marié, célibataire...**

Méthode :

- **Analyse Factorielle des Correspondances**

3- Les méthodes

Mesure
ordinaire

Les chiffres qui identifient la relation d'ordre entre les propriétés d'objet sans aucune relation de distance

Exemple :

- **Classement des goûts des clients selon un critère**
- **classe d'âge (15- 25), (26- 35)....**
- **le rang**

Méthode :

- **Analyse Factorielle des Correspondances**

3- Les méthodes

Mesure
ordinaire

Variable quantitative dont les valeurs ont une relation d'ordre et de distance

Exemple :

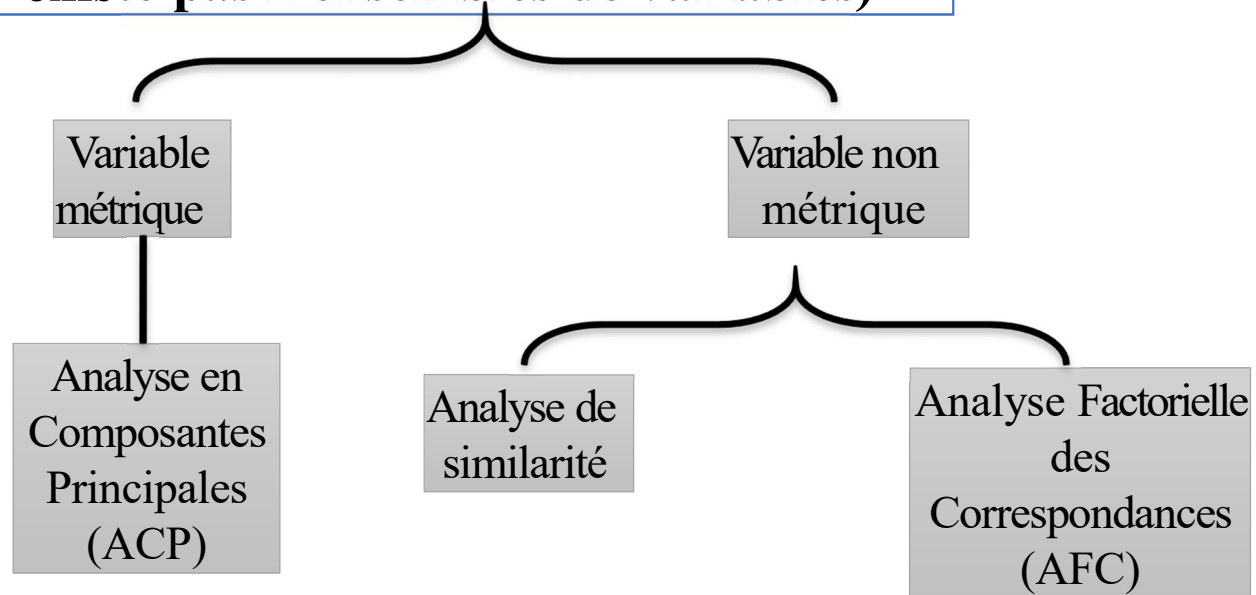
- **Le nombre de points de vente d'une marque**
- **La valeur d'investissement dans une ville**

Méthode :

- **Analyse en Composantes Principales**

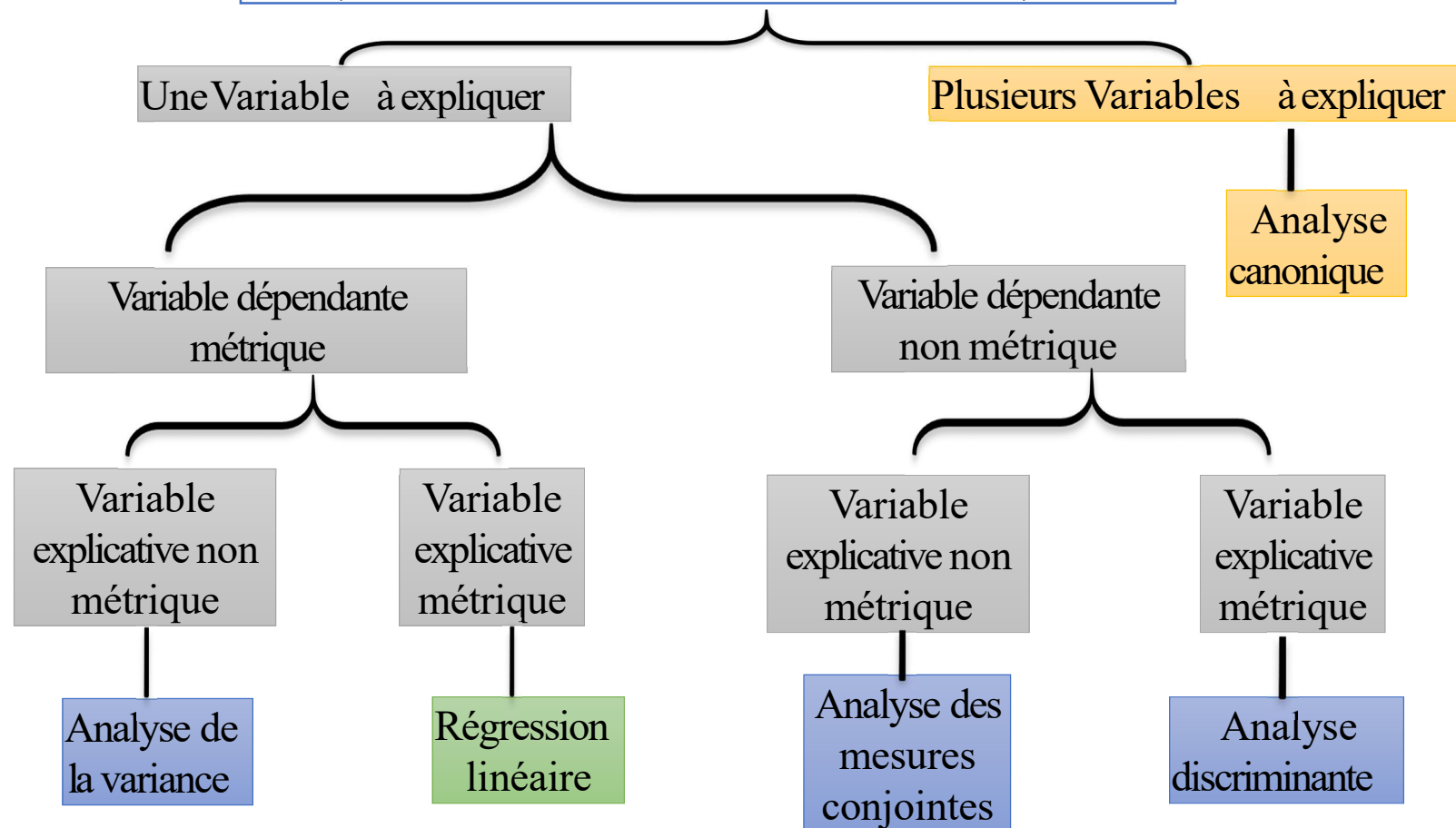
Méthodes descriptives

(N'existe pas 2 ensembles de variables)



Méthodes explicatives

(Existe 2 ensembles de variables)



Chapitre 2 : ACP

Introduction

Chapitre 1 : SVD

Chapitre 2 : Analyse en Composantes Principales

- 1. Définition**
- 2. Principe**
- 3. Démarche**
- 4. Cas d'application**

Chapitre 3 : Analyse Factorielle des Correspondances

Chapitre 4 : AFC Multiples (AFCM)

Chapitre 5 : Classification Ascendante Hiérarchique

Conclusion

Chapitre 2: AFC

Introduction

Chapitre 1 : SVD

Chapitre 2 : Analyse en Composantes Principales

Chapitre 3 : Analyse Factorielle des Correspondances

1. Définition

2. Objectifs

3. Démarche

4. Cas d'application

Chapitre 4 : AFC Multiples (AFCM)

Chapitre 5 : Classification Ascendante Hiérarchique

Conclusion

Chapitre 3: CAH

Introduction

Chapitre 1 : SVD

Chapitre 2 : Analyse en Composantes Principales

Chapitre 3 : Analyse Factorielle des Correspondances

Chapitre 4 : AFC Multiples (AFCM)

Chapitre 5 : Classification Ascendante Hiérarchique

1. Définition
2. La classification non hiérarchique
3. La classification hiérarchique
4. Cas d'application

Conclusion

Plan

Introduction

Chapitre 1 : SVD

Chapitre 2 : Analyse en Composantes Principales

Chapitre 3 : Analyse Factorielle des Correspondances

Chapitre 5 : AFC Multiples (AFCM)

Chapitre 6 : Classification Ascendante Hiérarchique

Conclusion