

CERTIFICAT DE SPÉCIALISATION ANALYSTE DE DONNÉES MASSIVES

Entreposage et fouille de données

Projet de STA211

Auteurs:

Boukary OUEDRAOGO

Bangaly CAMARA

Imad EL HAMMA

Professeur:

Mme Niang

2021-2022

 $U(x) \approx U_h = \sum_{j=1}^{N} c_j \varphi_j(x)$

$$u''(x_i) \approx \frac{1}{h^2} [u(x_{i+1}) - 2u(x_i) + u(x_{i-1})]$$

Table des matières

1	Intro	oductio	n .	1					
2	Cha	Chargement des données et analyses préliminaires							
	2.1	Chargement des données							
	2.2	Analys	e descriptive des variables quantitatives	1					
		2.2.1	Analyse univarié	1					
		2.2.2	ACP sur les variables quantitatives du bloc ménage	2					
		2.2.3	ACM sur les variables qualitatives du bloc ménage	2					
		2.2.4	AFM sur les variable du bloc ménage	3					
		2.2.5	ACP sur les variables quantitatives du bloc ménage	3					
		2.2.6	ACM sur les variables qualitatives du bloc ménage	4					
		2.2.7	AFM sur les variable du bloc ménage	5					
		2.2.8	ACP sur les variables quantitatives du bloc ménage	5					
		2.2.9	ACM sur les variables qualitatives du bloc ménage	5					
		2.2.10	AFM sur les variable du bloc habitude	6					

Résumé

1. Introduction

2. Chargement des données et analyses préliminaires

Cette partie est consacrée au chargement des données et à la sélection des variables liées au ménage. Certaines variables qui sont codées comme des variables numériques mais qui en réalité sont qualitatives seront recordées en variables facteurs.

2.1 Chargement des données

Après le chargement des données, l'étape suivante est l'analyse univariée. On peut regarder les statistiques descriptives simples avec la function **summary** et la fonction **describe**.

Description of .

2.2 Analyse descriptive des variables quantitatives

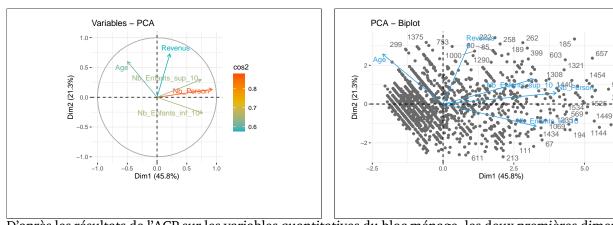
2.2.1 Analyse univarié

Variables	Moyenne	Médiane	Min	Maximum	Variance	Ecart-type
Age	51.87	52	18	89	209.91	14.49
Revenus	2 573.20	2 349	535	7 600	1 967 644.23	1 402.73
Nombre d'enfants dont âge > 10 ans	0.44	0	0	3	0.67	0.82
Nombre d'enfants dont âge <= 10 ans	0.32	0	0	3	0.48	0.70

TABLE 1 – Statistiques descriptives des variables quantitatives

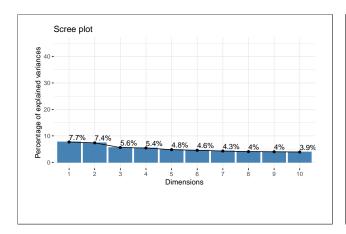
Etant donné la structure en groupe des données, nous allons réaliser des analyses factorielles sur chaque groupe de variables. Pour le bloc chaque bloc(ménage, habitude, logement), nous réalisons des **ACP** sur les variables quantitatives, une **ACM** sur les variables qualitatives et une **AFM** sur chaque bloc. L'objectif est de voir s'il existe une différence entre individus par rapport à chaque groupe de variable ou s'il existe des groupes d'individus homogènes par rapport à chaque groupe ou chaque bloc de variables.

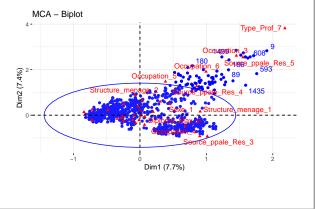
2.2.2 ACP sur les variables quantitatives du bloc ménage

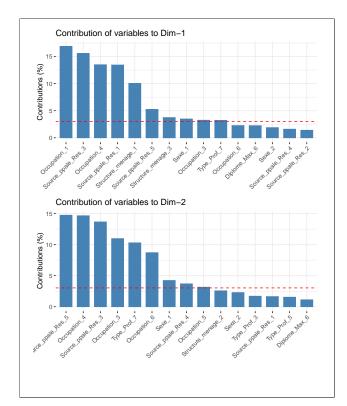


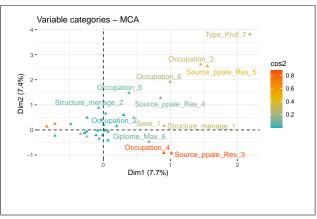
D'après les résultats de l'ACP sur les variables quantitatives du bloc ménage, les deux premières dimensions portent 61.7 de la variabilité contenu dans ce groupe de variables. Cela répresente un part significative. Les variables NBPerson,NBEnftssup0 et NBEnftsinf10 sont fortement corrélés au premier axe factoriel de l'ACP. A l'opposé, les variables Age et Revenus sont fortement corrélées au deuxième axe factoriel de l'ACP. Les individus(ménages) qui ont de fortes coordonnées sur l'axe 2 sont des ménages relativement aisés et âgés. Ceux qui ont de fortes coordonnées sur l'axe 1 sont des ménages composés de beaucoup de personnes et avec des enfants.

2.2.3 ACM sur les variables qualitatives du bloc ménage

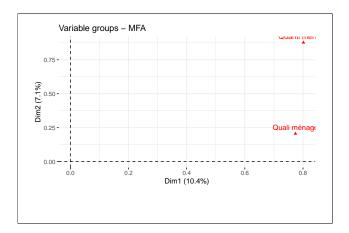




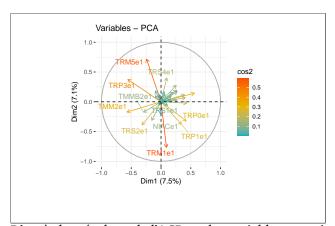


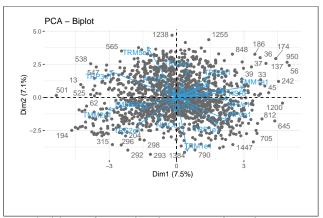


2.2.4 AFM sur les variable du bloc ménage



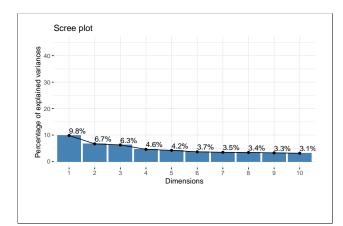
2.2.5 ACP sur les variables quantitatives du bloc ménage

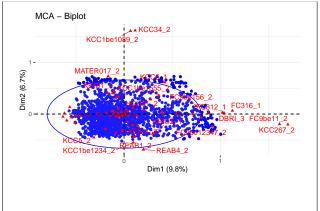


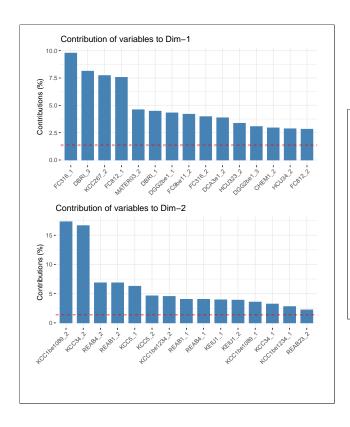


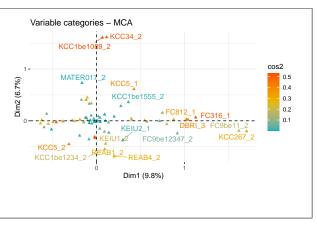
D'après les résultats de l'ACP sur les variables quantitatives du bloc ménage, les deux premières dimensions portent 61.7 de la variabilité contenu dans ce groupe de variables. Cela répresente un part significative. Les variables NBPerson,NBEnftssup0 et NBEnftsinf10 sont fortement corrélés au premier axe factoriel de l'ACP. A l'opposé, les variables Age et Revenus sont fortement corrélées au deuxième axe factoriel de l'ACP. Les individus(ménages) qui ont de fortes coordonnées sur l'axe 2 sont des ménages relativement aisés et âgés. Ceux qui ont de fortes coordonnées sur l'axe 1 sont des ménages composés de beaucoup de personnes et avec des enfants.

2.2.6 ACM sur les variables qualitatives du bloc ménage

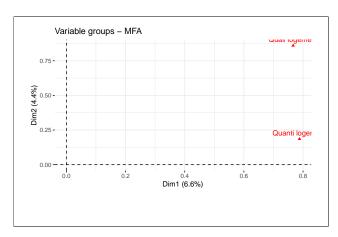




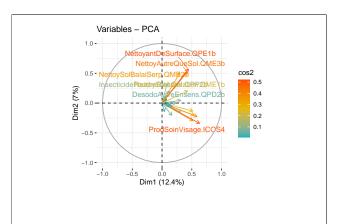


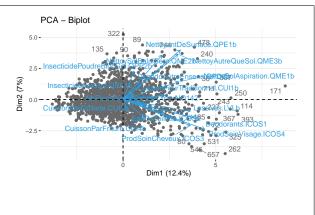


2.2.7 AFM sur les variable du bloc ménage



2.2.8 ACP sur les variables quantitatives du bloc ménage

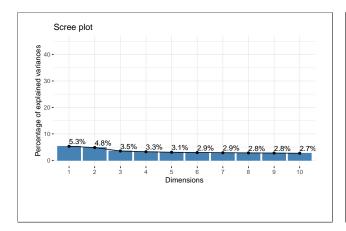


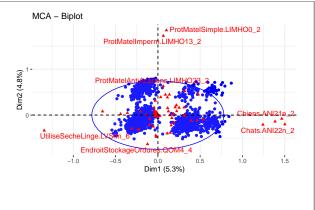


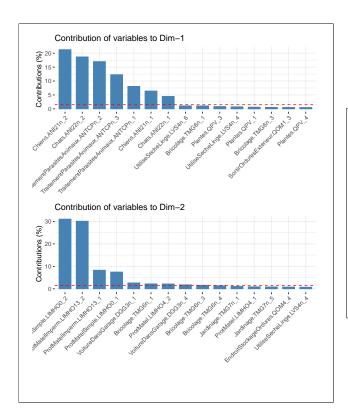
D'après les résultats de l'ACP sur les variables quantitatives du bloc ménage, les deux premières dimensions portent 61.7 de la variabilité contenu dans ce groupe de variables. Cela répresente un part significative.

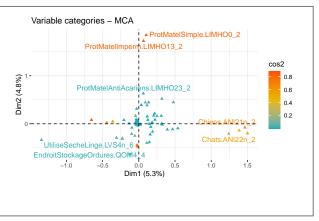
Les variables NBPerson, NBEnftssup0 et NBEnftsinf10 sont fortement corrélés au premier axe factoriel de l'ACP. A l'opposé, les variables Age et Revenus sont fortement corrélées au deuxième axe factoriel de l'ACP. Les individus (ménages) qui ont de fortes coordonnées sur l'axe 2 sont des ménages relativement aisés et âgés. Ceux qui ont de fortes coordonnées sur l'axe 1 sont des ménages composés de beaucoup de personnes et avec des enfants.

2.2.9 ACM sur les variables qualitatives du bloc ménage









2.2.10 AFM sur les variable du bloc habitude

