

---

# 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

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Analyse exploratoire</b>	<b>2</b>
2.1	Analyse univariée . . . . .	2
2.2	Analyse bivariée . . . . .	3
2.3	Analyse exploration multidimensionnelle . . . . .	3
2.3.1	Analyse en composante principale(ACP) . . . . .	3
2.3.2	Analyse des correspondances multiples(ACM) . . . . .	3
2.3.3	Analyse factorielle multiple(AFM) . . . . .	3
2.3.4	AFM sur l'ensemble des variables du jeu de données . . . . .	3
2.3.5	Classification des individus . . . . .	3
2.3.6	Classification des variables . . . . .	3
<b>3</b>	<b>Modélisation et prédiction du Formaldéhyde</b>	<b>3</b>
3.1	Traitement des données manquantes . . . . .	3
3.2	Modélisation . . . . .	3
3.2.1	Random Forest . . . . .	3
3.2.2	Fuzzy forest sur les blocs initiaux . . . . .	3
3.2.3	Fuzzy forest sur sur les groupes issues de ClusterOfvar . . . . .	3
3.2.4	Comparaison et selection de modèle . . . . .	3
<b>4</b>	<b>Conclusion</b>	<b>3</b>
4.0.1	ACP sur les variables quantitatives du bloc ménage . . . . .	3
4.0.2	ACM sur les variables qualitatives du bloc ménage . . . . .	4
4.0.3	AFM sur les variable du bloc ménage . . . . .	4
4.0.4	ACP sur les variables quantitatives du bloc ménage . . . . .	4
4.0.5	ACM sur les variables qualitatives du bloc ménage . . . . .	5
4.0.6	AFM sur les variable du bloc ménage . . . . .	5
4.0.7	ACP sur les variables quantitatives du bloc ménage . . . . .	5
4.0.8	ACM sur les variables qualitatives du bloc ménage . . . . .	6
4.0.9	AFM sur les variable du bloc habitude . . . . .	6

## Résumé

## 1. Introduction

## 2. Analyse exploratoire

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 Analyse univariée

## Description of .

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

## 2.2 Analyse bivariée

## 2.3 Analyse exploration multidimensionnelle

### 2.3.1 Analyse en composante principale(ACP)

### 2.3.2 Analyse des correspondances multiples(ACM)

### 2.3.3 Analyse factorielle multiple(AFM)

### 2.3.4 AFM sur l'ensemble des variables du jeu de données

### 2.3.5 Classification des individus

### 2.3.6 Classification des variables

## 3. Modélisation et prédiction du Formaldéhyde

### 3.1 Traitement des données manquantes

### 3.2 Modélisation

#### 3.2.1 Random Forest

#### 3.2.2 Fuzzy forest sur les blocs initiaux

#### 3.2.3 Fuzzy forest sur sur les groupes issues de ClusterOfvar

#### 3.2.4 Comparaison et selection de modèle

## 4. Conclusion

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.

### 4.0.1 ACP sur les variables quantitatives du bloc ménage

```
## Error in .get_facto_class(X):  
objet 'res.pca_menage' introuvable
```

```
## Error in .get_facto_class(X):  
objet 'res.pca_menage' introuvable
```

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 représente 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.

#### 4.0.2 ACM sur les variables qualitatives du bloc ménage

```
## Error in get_eig(X): objet
'res.mca_menage' introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_menage' introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_menage' introuvable
## Error in .get_facto_class(X):
objet 'res.mca_menage' introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_menage' introuvable
```

#### 4.0.3 AFM sur les variable du bloc ménage

```
## Error in .get_facto_class(X):
objet 'res.mfa_menage' introuvable
```

#### 4.0.4 ACP sur les variables quantitatives du bloc ménage

```
## Error in .get_facto_class(X):
objet 'res.pca_logement'
introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.pca_logement'
introuvable
```

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 représente 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.

#### 4.0.5 ACM sur les variables qualitatives du bloc ménage

```
## Error in get_eig(X): objet  
'res.mca_logement' introuvable
```

```
## Error in .get_facto_class(X):  
objet 'res.mca_logement'  
introuvable
```

```
## Error in .get_facto_class(X):  
objet 'res.mca_logement'  
introuvable  
## Error in .get_facto_class(X):  
objet 'res.mca_logement'  
introuvable
```

```
## Error in .get_facto_class(X):  
objet 'res.mca_logement'  
introuvable
```

#### 4.0.6 AFM sur les variable du bloc ménage

```
## Error in .get_facto_class(X):  
objet 'res.mfa_logement'  
introuvable
```

#### 4.0.7 ACP sur les variables quantitatives du bloc ménage

```
## Error in .get_facto_class(X):  
objet 'res.pca_habitude'  
introuvable
```

```
## Error in .get_facto_class(X):  
objet 'res.pca_habitude'  
introuvable
```

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 représente 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.

#### 4.0.8 ACM sur les variables qualitatives du bloc ménage

```
## Error in get_eig(X): objet
'res.mca_habitude' introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_habitude'
introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_habitude'
introuvable
## Error in .get_facto_class(X):
objet 'res.mca_habitude'
introuvable
```

```
## Error in .get_facto_class(X):
objet 'res.mca_habitude'
introuvable
```

#### 4.0.9 AFM sur les variable du bloc habitude

```
## Error in .get_facto_class(X):
objet 'res.mfa_logement'
introuvable
```

```
## Error in kable_latex(x = structure(c("Name", "Number of rows", "Number of columns",
: argument inutilisé (table.attr = "style='width: auto;'\n          class='table table-cond
```

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