|  |
| --- |
| **REPUBLIQUE DU SENEGAL** |

|  |
| --- |
|  |

|  |
| --- |
| **\*\*\*\*\*\*\*\*\*\*** |
| ***Un Peuple - Un But - Une Foi*** |
| **\*\*\*\*\*\*\*\*\*\*** |
| **Agence nationale de la Statistique et de la démographie** |

|  |
| --- |
|  |

|  |
| --- |
| **\*\*\*\*\*\*\*\*\*\*** |
| **Ecole nationale de la Statistique et de l'Analyse économique Pierre Ndiaye** |

|  |
| --- |
|  |

# EVALUATION

|  |
| --- |
| *Projet statistique avec R* |

|  |  |
| --- | --- |
| **Rédigé par** | **Sous la supervision de** |
| Mame Balla BOUSSO | M. Aboubacar HEMA |
| *Élève Ingénieur Statisticien Économiste* | *Analyste de recherche* |

|  |
| --- |
| **Année académique 2024-2025** |

# Sommaire

## Liste des figures

## Liste des tableaux

# I. Analyse de consistence des bases de données:

#### Aperçu des deux bases

[1] 8950 100

[1] 2206 26

La base de donnée Base\_Principale contient 8950 observations et 100 variables, alors celle MAD contient 2206 lignes et 26 variables.

#### Analyses des deux bases

Par ailleurs l’application de clean\_names() du package a permis de nettoyer les noms des colonnes. La fonction names offre cet appercu dans les noms des colonnes

Cette action permet d’uniformiser les variables dans les deux bases. Elle peut trouver son utilité notamment lorqu’on souhaite merger des bases. Elle permet de rendre uniforme à l’occurrence les clés.

#### Vérification de doublons sur l’indentification des ménages

Nous avons examiné la variable id dans chaque base afin d’identifier d’éventuels doublons :

Dans la **Base\_Principale**, chaque identifiant est unique. Aucune duplication n’a été détectée (n = 0 pour tous les id).

En revanche, **Base\_MAD** contient plusieurs doublons d’identifiants. Par exemple, l’identifiant ‘1050’ est présent 3 fois. De plus, 9 lignes ont un identifiant vide (’’)

#### Analyse des valeurs manquantes

| variable | pct\_na |
| --- | --- |
| hhh\_main\_activity | 100.00000 |
| hhh\_matrimonial | 100.00000 |
| fcs\_pr\_s\_rf | 100.00000 |
| fcs\_fruit\_s\_rf | 96.94972 |
| fcs\_veg\_s\_rf | 88.69274 |
| fcs\_fruit\_org | 85.88827 |
| hh\_source\_income | 80.87151 |
| hdds\_ch | 63.22905 |
| hdds\_fruit\_org | 61.40782 |
| hdds\_pr\_egg | 59.81006 |

On voit ainsi le aux de valeurs manquantes pour les 10 variables avec le plus de valeurs manquantes. On voit la plupart des variables présentent un très grand nombre de valeurs manquantes. On va supprimer les varibales ayant un taux de NA supérieur à 40%.

Ce chois peut-être justifié par le fait que :

* au-delà de ce seuil, les estimations par imputation deviennent peu fiables ;
* l’analyse peut être biaisée ou affaiblie si trop de données sont absentes.

Variables supprimées (taux de NA > 40 %):

[1] "hhh\_main\_activity" "hhh\_matrimonial" "hh\_source\_income"   
 [4] "hdds\_ch" "hdds\_veg\_org" "hdds\_fruit\_org"   
 [7] "hdds\_fruit\_oth" "hdds\_pr\_meat\_o" "hdds\_pr\_egg"   
[10] "fcs\_dairy\_s\_rf" "fcs\_pr\_s\_rf" "fcs\_veg\_s\_rf"   
[13] "fcs\_fruit\_s\_rf" "fcs\_fruit\_org"

| variable | pct\_na |
| --- | --- |
| ever\_breast\_f | 0.3173164 |

Par contre dans la base, **BASE\_MAD**, on a une seule variable ayant des NA et le taux est très faible.

#### Vérification de la cohérence dans la variables age

Dans l’analyse que l’on s’apprète à faire,la variables âge sera déterminante. Par conséquent, il serait important de vérifier sa validité.

On vérifie la cohérence entre la taille totale du ménage (hh\_size) et la somme des membres du ménage par tranche d’âge et sexe. Le résultat montre une parfaite logique.

| prop\_coherent |
| --- |
| 1 |

En effet, aa proportion prop\_coherent est égale à 1. Cela signifie que dans tous les cas de la base\_principale, la somme des tailles de ménage par tranche d’âge et sexe est cohérente avec la taille totale du ménage. Aucun ménage ne présente d’incohérence dans ce calcul.

# II. Analyse des données et calcul d’indicateurs

## 1) Faites une analyse socio-démographique des ménages

| **Characteristic** | **N = 8,950***1* |
| --- | --- |
| Sexe du chef de ménage |  |
| Femme | 3,938 (44%) |
| Homme | 5,012 (56%) |
| Âge du chef de ménage | 43 (13) |
| Taille du ménage | 7.55 (12.90) |
| Niveau d’éducation du chef de ménage |  |
| Aucune | 1,827 (31%) |
| Alphabétisé ou Coranique | 3,684 (62%) |
| Primaire | 265 (4.5%) |
| Secondaire | 126 (2.1%) |
| Superieur | 24 (0.4%) |
| Unknown | 3,024 |
| *1*n (%); Mean (SD) | |

Le tableau présente des caractéristiques sociodémographiques des ménages. Les chefs de ménage sont majoritairement des hommes (56 %, soit 5 012 individus), contre 44 % de femmes (3 938 individus), reflétant des normes culturelles où les hommes sont plus souvent désignés comme chefs de famille. L’âge moyen du chef de ménage est de 43 ans (avec un écart-type de 13 ans), indiquant une population en âge d’assumer des responsabilités économiques et familiales. La taille moyenne des ménages est de 7,55 personnes, avec un écart-type élevé (12,90), suggérant une grande variabilité, allant de petits noyaux familiaux à des familles étendues nombreuses, typique des sociétés rurales.

En matière d’éducation, les chefs de ménage ont principalement un faible niveau de formation : 31 % n’ont aucune éducation (1 827 individus), tandis que 62 % (3 684) sont alphabétisés ou ont une éducation coranique. Seulement 4,5 % ont un niveau primaire, 2,1 % secondaire, et 0,4 % supérieur, soulignant des obstacles structurels à l’accès à l’éducation formelle. Cependant, 3 024 cas n’ont pas de données renseignées, ce qui limite l’interprétation.

## 2) Score de consommation alimentaire (SCA)

### a) Analyse descriptive des variables qui composent le SCA

L’analyse des variables composant le Score de Consommation Alimentaire (SCA) repose sur la fréquence de consommation de différents groupes alimentaires au sein des ménages, mesurée sur une période de 7 jours. Cinq principaux groupes alimentaires ont été identifiés, chacun se verra attribuer un poids attribué en fonction de son importance nutritionnelle dans le calcul du SCA.

Le premier groupe, les céréales, racines et tubercules (représenté par la variable fcs\_stap), inclut des aliments essentiels à l’alimentation de base dans de nombreuses régions. Le deuxième groupe est constitué des légumineuses (variable fcs\_pulse), qui comprennent des aliments comme les haricots, les pois et les lentilles.

Les produits laitiers, représentés par la variable fcs\_dairy, comprennent des produits tels que le lait, le fromage et le yaourt.

Le groupe suivant regroupe les viandes, poissons et œufs (combinant les variables fcs\_pr\_meat\_f, fcs\_pr\_meat\_o, fcs\_pr\_fish et fcs\_pr\_egg), qui représentent des sources majeures de protéines animales et d’acides gras essentiels.

Enfin, les légumes (variable fcs\_veg), qui incluent aussi bien les légumes frais que cuits.

Le calcul du Score de Consommation Alimentaire (SCA) consiste à multiplier la fréquence de consommation de chaque groupe alimentaire par son poids respectif. La somme de ces valeurs pondérées donne un score global, qui reflète la diversité alimentaire et l’équilibre nutritionnel du ménage au cours des sept derniers jours. Ce score permet de mieux comprendre les habitudes alimentaires des ménages et de les évaluer par rapport à des standards de sécurité alimentaire et de nutrition.

#### Attribution des poids

Les poids attribués à chaque groupe alimentaire dans le calcul du Score de Consommation Alimentaire (SCA) sont basés sur leur importance nutritionnelle et leur contribution à un régime alimentaire équilibré. En accord avec les recommandations de l’OMS :

* Les céréales sont la base énergétique de l’alimentation (Poids = 3).
* Les légumineuses apportent une alternative de qualité pour les protéines végétales (Poids = 2).
* Les produits laitiers jouent un rôle clé dans la santé osseuse et la fourniture de nutriments essentiels (Poids = 4).
* Les viandes, poissons et œufs sont essentiels pour des protéines de haute qualité et des micronutriments cruciaux (Poids = 5).
* Les légumes fournissent des micronutriments et des fibres importantes pour la prévention des maladies (Poids = 2).

Les poids attribués à chaque groupe alimentaire dans le SCA sont donc un reflet de leur rôle dans une alimentation saine et équilibrée, en mettant l’accent sur la diversité des apports nutritionnels.

### b) Calculer le score de consommation alimentaire

Le calcul du SCA repose sur l’agrégation des variables mentionnées précédemment, avec les poids spécifiés.

Formule générale :

Où chaque variable (fcs\_\*) représente le nombre de jours de consommation de chaque catégorie alimentaire, et les poids sont appliqués pour refléter leur importance respective dans le score final.

On a aisin les valeurs de SCA

### c) Tableau illustrant le poids attribue a chaque groupe alimentaire pour le calcul du SCA (la somme totale des poids doit etre egale a 16)

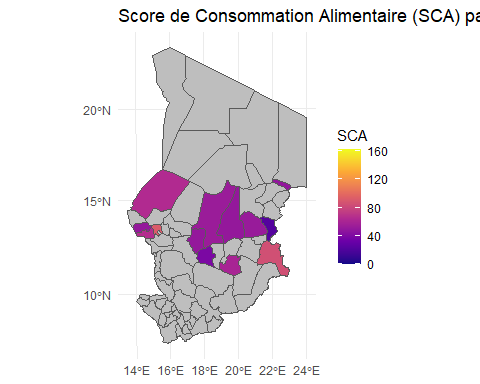
| Groupe alimentaire | Variables | Poids |
| --- | --- | --- |
| Céréales, racines, tubercules | fcs\_stap | 3 |
| Légumineuses | fcs\_pulse | 2 |
| Produits laitiers | fcs\_dairy | 4 |
| Viandes, poisson, œufs | fcs\_pr\_meat\_f + fcs\_pr\_meat\_o + fcs\_pr\_fish + fcs\_pr\_egg | 5 |
| Légumes | fcs\_veg | 2 |

On fait ensuit la classification

Reading layer `tcd\_admbnda\_adm1\_20250212\_AB' from data source   
 `C:\Users\lenovo\Desktop\Semestre\_6\Projet-statistique-sous-R\RENDUS\_EXAMEN\_ISEP3\Mame Balla BOUSSO\Data\shapefile\tcd\_admbnda\_adm1\_20250212\_AB.shp'   
 using driver `ESRI Shapefile'  
Simple feature collection with 23 features and 4 fields  
Geometry type: POLYGON  
Dimension: XY  
Bounding box: xmin: -1203504 ymin: 878475.6 xmax: -103563.6 ymax: 2693715  
Projected CRS: Africa\_Albers\_Equal\_Area\_Conic

Reading layer `tcd\_admbnda\_adm2\_20250212\_AB' from data source   
 `C:\Users\lenovo\Desktop\Semestre\_6\Projet-statistique-sous-R\RENDUS\_EXAMEN\_ISEP3\Mame Balla BOUSSO\Data\shapefile\tcd\_admbnda\_adm2\_20250212\_AB.shp'   
 using driver `ESRI Shapefile'  
Simple feature collection with 70 features and 7 fields  
Geometry type: POLYGON  
Dimension: XY  
Bounding box: xmin: -1203504 ymin: 878475.6 xmax: -103563.6 ymax: 2693715  
Projected CRS: Africa\_Albers\_Equal\_Area\_Conic

#### SCA par département



## 3) L’indice réduit des stratégies de survie (rCSI):

L’indice réduit des stratégies de survie, basé sur la consommation, est utilisé pour évaluer le niveau de stress d’un ménage dû à une pénurie alimentaire. Il mesure les stratégies comportementales que les gens appliquent lorsqu’ils n’ont pas accès à suffisamment de nourriture ou lorsqu’ils prévoient une diminution de la sécurité alimentaire.

### a) Analyse descriptive des variables qui composent le rCSI

Nous allons donner le tableau qui suit pour servir de descriotion à nos variables du rCSI.

| **Characteristic** | **N = 8,950***1* |
| --- | --- |
| **Sold household assets/goods (radio, furniture, refrigerator, television, jewelle** |  |
| No, because I did not need to | 0 (0%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 0 (0%) |
| Yes | 0 (0%) |
| Not applicable (don't have children/these assets) | 8,950 (100%) |
| **Spent savings due to lack of food** |  |
| No, because I did not need to | 3,156 (35%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,473 (16%) |
| Yes | 420 (4.7%) |
| Not applicable (don't have children/these assets) | 3,901 (44%) |
| **Sent household members to eat elsewhere/live with family or friends due to lack** |  |
| No, because I did not need to | 2,763 (31%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,233 (14%) |
| Yes | 375 (4.2%) |
| Not applicable (don't have children/these assets) | 4,579 (51%) |
| **Purchased food/non-food on credit (incur debts) due to lack of food** |  |
| No, because I did not need to | 3,258 (36%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,384 (15%) |
| Yes | 1,719 (19%) |
| Not applicable (don't have children/these assets) | 2,589 (29%) |
| **Sold productive assets or means of transport (sewing machine, wheelbarrow, bicyc** |  |
| No, because I did not need to | 3,487 (39%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,549 (17%) |
| Yes | 300 (3.4%) |
| Not applicable (don't have children/these assets) | 3,614 (40%) |
| **Reduced expenses on health (including drugs)** |  |
| No, because I did not need to | 2,870 (32%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,269 (14%) |
| Yes | 163 (1.8%) |
| Not applicable (don't have children/these assets) | 4,648 (52%) |
| **Withdrew children from school due to lack of food** |  |
| No, because I did not need to | 3,231 (36%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,397 (16%) |
| Yes | 108 (1.2%) |
| Not applicable (don't have children/these assets) | 4,214 (47%) |
| **Mortgaged/Sold house or land due to lack of food** |  |
| No, because I did not need to | 3,259 (36%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,387 (15%) |
| Yes | 53 (0.6%) |
| Not applicable (don't have children/these assets) | 4,251 (47%) |
| **Begged and/or scavenged (asked strangers for money/food) due to lack of food** |  |
| No, because I did not need to | 3,140 (35%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,437 (16%) |
| Yes | 282 (3.2%) |
| Not applicable (don't have children/these assets) | 4,091 (46%) |
| **Engaged in illegal income activities (theft, prostitution) due to lack of food** |  |
| No, because I did not need to | 3,138 (35%) |
| No, because I already sold those assets or have engaged in this activity within the last 12 months and cannot continue to do it | 1,379 (15%) |
| Yes | 202 (2.3%) |
| Not applicable (don't have children/these assets) | 4,231 (47%) |
| *1*n (%) | |

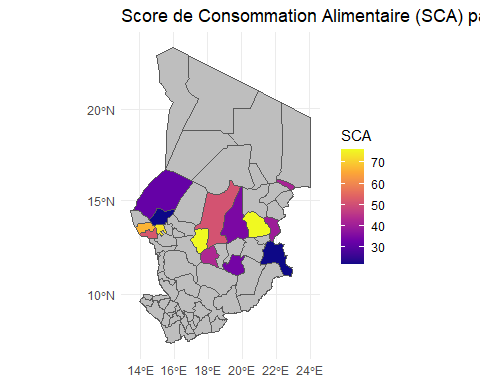
### b) Calculer l’indice réduit des stratégies de survie

### c) Rableau illustrant le poids attribue a chaque variable pour le calcul du rCSI (la somme totale des poids doit etre egale a 21)

Poids attribués aux variables pour le calcul du rCSI (somme = 21)

| **Stratégie** | **Variable** | **Poids** |
| --- | --- | --- |
| Stress1 | lh\_csi\_stress1 | 1 |
| Stress2 | lh\_csi\_stress2 | 1 |
| Stress3 | lh\_csi\_stress3 | 1 |
| Stress4 | lh\_csi\_stress4 | 1 |
| Crisis1 | lh\_csi\_crisis1 | 2 |
| Crisis2 | lh\_csi\_crisis2 | 2 |
| Crisis3 | lh\_csi\_crisis3 | 2 |
| Emergency1 | lh\_csi\_emergency1 | 3 |
| Emergency2 | lh\_csi\_emergency2 | 3 |
| Emergency3 | lh\_csi\_emergency3 | 3 |

### d) Faites une representation spatiale (region et departement) du rCSI

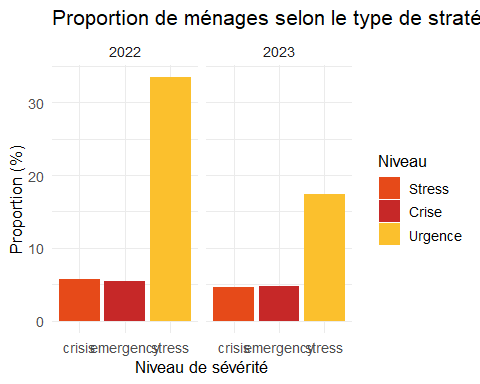


## 4) Stratégies d’adaptation aux moyens d’existence (LhCSI)

Les indicateurs de stratégies d’adaptation des moyens de subsistance (LCS) sont des indicateurs mesurés au niveau des ménages qui sont relativement simples et qui sont en corrélation avec d’autres mesures de la sécurité alimentaire et de la vulnérabilité. Ils sont basés sur une série de questions concernant la manière dont les ménages parviennent à faire face aux chocs qui mettent à mal leurs moyens de subsistance. Les indicateurs LCS évaluent les capacités d’adaptation et de production des ménages à moyen et à long terme et leur impact futur sur les besoins alimentaires. L’indicateur des stratégies d’adaptation des moyens de subsistance existe en deux versions. L’une pour la sécurité alimentaire (LCS-FS) et l’autre pour les besoins essentiels (LCS-EN). Pour le cadre de l’examen on s’interesse a LCS-FS.

### a) Faites une analyse descriptive des variables qui composent le LhCSI.

### b) La proportion de menage en situation de stress, de crise et d’urgence en 2022 et 2023



### c) Representation spatiale (region et departement) des strategies d’adaptation

## 5) Score de diversité alimentaire des ménages

### a) Analyse descriptive des variables qui composent le module HDDS

### b) Le score de diversité alimentaire des ménages c’est a dire le nombre total de groupes alimentaires consommés par les membres du ménage.

### c) Representation spatiale (region et departement) du score de diversité alimentaire

# Table des matières