|  |
| --- |
| **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** |
|  |

##### Projet statistique sur R : Evaluation

|  |  |
| --- | --- |
| **Rédigé par** | **Sous la supervision de** |
| LAWA FOUMSOU Prosper | M. Aboubacar HEMA |
| *Élève Ingénieur Statisticien Économiste* | *Data-scientist* |

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

# Introduction

Cette étude s’inscrit dans le cadre de l’**Enquête sur les déplacements forcés au Sud-Soudan menée en 2023**, un pays affecté par des crises humanitaires prolongées dues à des conflits armés, des catastrophes naturelles et des déplacements massifs de population. Ces dynamiques ont eu un impact significatif sur la sécurité alimentaire des ménages, en particulier parmi les populations déplacées internes (IDPs).

L’objectif principal de cette analyse est d’évaluer la qualité de la consommation alimentaire des ménages à l’aide du **Score de Consommation Alimentaire (SCA)**, un indicateur développé par le **Programme Alimentaire Mondial (PAM)**. Ce score combine la diversité alimentaire, la fréquence de consommation et la valeur nutritionnelle des groupes alimentaires consommés durant les 7 derniers jours.

L’utilisation de **R** permet d’effectuer un traitement reproductible des données, de calculer le SCA selon les normes du PAM, et de classifier les ménages en fonction de leur niveau de sécurité alimentaire. Cette analyse repose sur les seuils adaptés au contexte du **Sud-Soudan**, en tenant compte des recommandations techniques du PAM en matière de classification.

Les résultats obtenus permettront d’identifier les ménages les plus vulnérables en matière d’accès à une alimentation adéquate et d’orienter les interventions humanitaires de manière ciblée et efficace.

# I. Importation et Analyse de consistance des bases

## 1. Importation des jeux de données

Commençons par avoir une idée du nombre de variables et d’observations dans chaque base.

### Base principale

La base principale a 3058 lignes et 1312 colonnes.

### Base MAD

La base individus a 22092 lignes et 22092 colonnes.

# II. Analyse de consistance des bases

La base principale a 3058 lignes et 1244 colonnes après suppression de colonnes vides.La base individus a 22092 lignes et 22092 colonnes après suppression de colonnes vides.

## Detection de doublons

On remarque qi’il n’ y a pas de doublons dans la base principale

Aussi on remarque qu’il n’y a pas de doublons dans la base individus. Passons donc au traitement des données manquantes

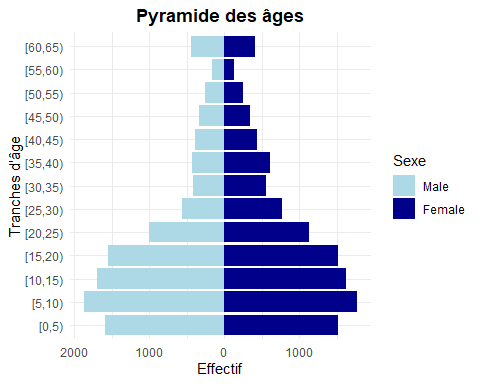
## Traitement de données manquantes

Cette partie servira à traiter les bases que nous utiliserons pour le travail. On procèdera notamment au traitement des doublons et des valeurs manquantes.

# Analyse socio-economique des deplaces internes du Sudan en 2023

## 1) Pyramide des ages des individus

Cette pyramide des âges représente la structure par âge et par sexe d’une population, divisée en tranches de 5 ans, allant de [0,5] à [60,65]. L’axe horizontal indique les effectifs, avec une échelle allant jusqu’à 2000 individus, tandis que l’axe vertical liste les groupes d’âge. Bien que le graphique ne montre pas explicitement les barres de distribution, la présence des labels “Male” et “Female” suggère une comparaison classique entre les sexes, avec les hommes traditionnellement représentés à gauche (valeurs négatives) et les femmes à droite (valeurs positives). La forme générale, si elle était visible, pourrait révéler si la population est jeune, vieillissante ou en transition démographique. Cependant, sans visualisation des données réelles, une analyse approfondie des tendances démographiques reste limitée.



Pyramide des âges par sexe

## 2) Caracteristiques du chef de menage

Le tableau présente les caractéristiques socio-démographiques des chefs de ménage parmi les réfugiés et les membres de la communauté hôte du nord. L’âge moyen des chefs de ménage est légèrement plus faible chez les réfugiés (27 ans, écart-type de 14) comparé à celui de la communauté hôte (29 ans, écart-type de 15), suggérant une population réfugiée légèrement plus jeune. La répartition par sexe est similaire dans les deux groupes, avec une légère majorité de femmes (52 %) à la tête des ménages. En ce qui concerne la situation matrimoniale, les différences sont plus marquées : les réfugiés sont plus fréquemment célibataires (54 % contre 47 % chez les hôtes), tandis que le mariage monogame est plus courant chez les membres de la communauté hôte (44 % contre 35 % chez les réfugiés). Les autres statuts matrimoniaux (polygamie, divorce, séparation, veuvage) restent minoritaires dans les deux groupes, bien que certaines variations existent. Ces résultats mettent en évidence des différences sociodémographiques notables entre les deux populations, qui peuvent refléter des dynamiques sociales et économiques spécifiques aux contextes de déplacement et d’accueil.

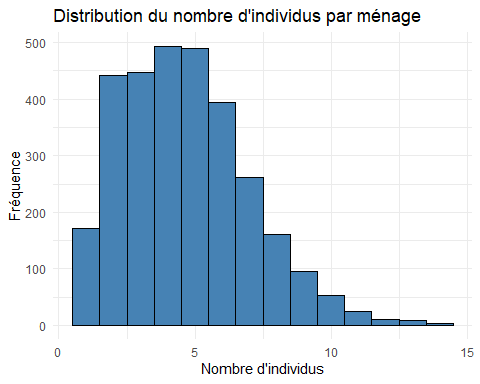
| **Variables socio-démographiques** | **N** | **Refugees**, N = 9,2851 | **Host community North**, N = 4,9631 |
| --- | --- | --- | --- |
| **Âge du chef de ménage** | 14,248 | 27 (14) | 29 (15) |
| **Sexe du chef de ménage** | 14,248 |  |  |
| *Male* |  | 4,417 (48%) | 2,405 (48%) |
| *Female* |  | 4,868 (52%) | 2,558 (52%) |
| **Situation matrimoniale** | 14,248 |  |  |
| *monogamous/married* |  | 3,217 (35%) | 2,176 (44%) |
| *polygamous/married* |  | 438 (4.7%) | 210 (4.2%) |
| *non-formal union* |  | 29 (0.3%) | 8 (0.2%) |
| *separated* |  | 207 (2.2%) | 53 (1.1%) |
| *divorced* |  | 127 (1.4%) | 30 (0.6%) |
| *widow or widower* |  | 256 (2.8%) | 165 (3.3%) |
| *never married* |  | 5,011 (54%) | 2,321 (47%) |
| 1Mean (SD); n (%) | | | |

## 3) Crowding Index ou l’indice d’affluence

## a) Calculez le nombre d’individu par menage

La taille des ménages dans l’échantillon varie de 1 à 14 personnes, avec une moyenne de 4,66 individus par ménage. La médiane est de 4 personnes, ce qui signifie que la moitié des ménages comptent au plus quatre membres. Le premier quartile (Q1) est de 3 et le troisième quartile (Q3) de 6, indiquant que 50 % des ménages ont entre 3 et 6 personnes. La moyenne étant légèrement supérieure à la médiane, cela suggère une légère asymétrie vers les valeurs élevées, due à la présence de quelques ménages très nombreux. Ces résultats traduisent une diversité dans la composition des ménages, avec une majorité de structures de taille moyenne, mais aussi l’existence de grandes unités familiales, potentiellement liées à des dynamiques culturelles, économiques ou contextuelles comme le déplacement forcé ou la cohabitation intergénérationnelle.

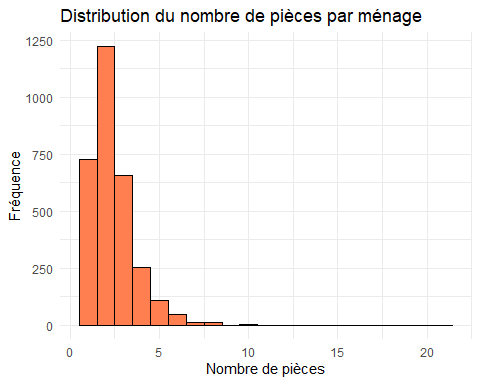
Min. 1st Qu. Median Mean 3rd Qu. Max.   
 1.000 3.000 4.000 4.664 6.000 14.000



## Faites des statistiques descriptives sur la variable renseigant le nombre de pieces du menage

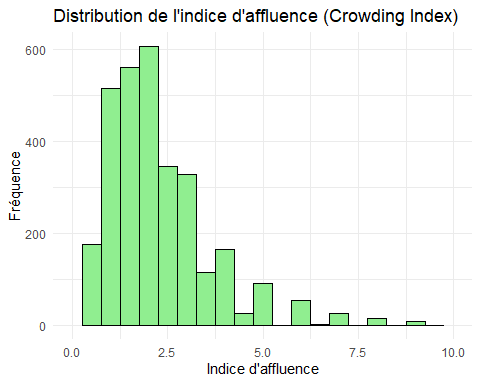
La variable analysée représente le **nombre de pièces** par logement. Elle varie entre **1 et 21 pièces**, avec une **moyenne de 2,39 pièces** par logement. La **médiane est de 2 pièces**, ce qui signifie que la moitié des logements disposent de deux pièces ou moins. Le **premier quartile (Q1)** est également de 2, tandis que le **troisième quartile (Q3)** est de 3, indiquant que 75 % des logements possèdent au plus trois pièces. La distribution est donc **fortement centrée sur les petites surfaces**, avec une large majorité de logements de 1 à 3 pièces. Toutefois, la valeur maximale de 21 pièces révèle la présence de logements nettement plus spacieux, bien que rares. La variable comporte également **6 valeurs manquantes**, qui devront être prises en compte dans l’analyse. Globalement, ces données suggèrent que les ménages vivent majoritairement dans des habitations de petite taille, possiblement en lien avec des contraintes économiques ou contextuelles.

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
 1.000 2.000 2.000 2.388 3.000 21.000 6

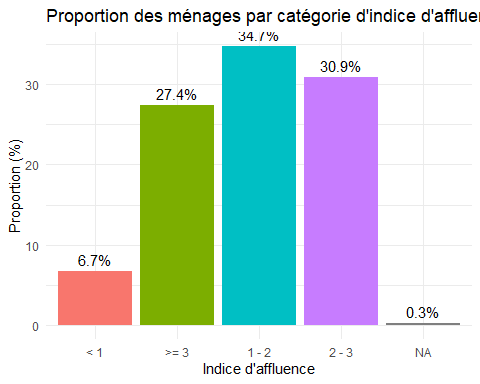


## d) Calculez l’indice d’affluence ou crowding index

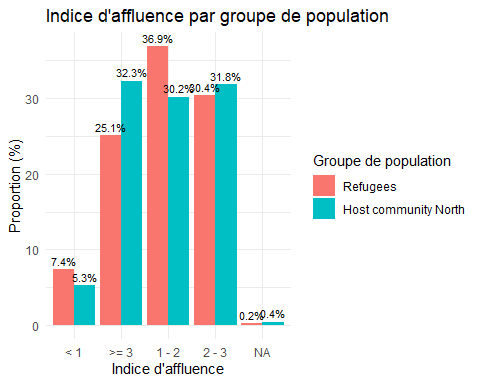
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
 0.125 1.333 2.000 2.278 3.000 10.000 9

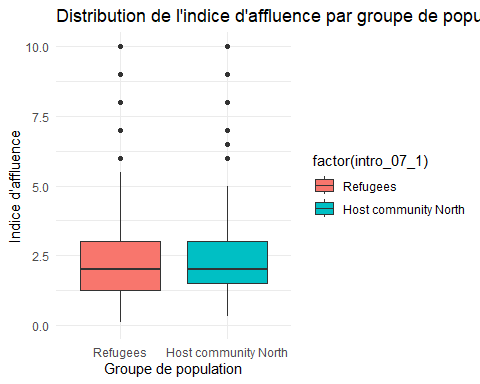


## e) Quelle est la proportion de menage dont le crowding index est: <1, 1<= crowding\_index<2, 2<= crowding\_index<3 et >=3



## Comparaison entre réfugiés et communautés d’accueil





# Analyse de la securite alimentaire des deplaces internes

## Score de consommation alimentaire (SCA)

### Faites une analyse descriptive des variables qui composent le SCA

Les variables concernées sont : food\_div1, food\_div2, food\_div3, food\_div4, food\_div5, food\_div6, food\_div7, food\_div8.

| **Characteristic** | **N = 3,0581** |
| --- | --- |
| Céréales & tubercules (jours) |  |
| 0 | 677 (22%) |
| 1 | 450 (15%) |
| 2 | 253 (8.3%) |
| 3 | 365 (12%) |
| 4 | 230 (7.5%) |
| 5 | 297 (9.7%) |
| 6 | 92 (3.0%) |
| 7 | 690 (23%) |
| Légumineuses/noix (jours) |  |
| 0 | 1,900 (62%) |
| 1 | 317 (10%) |
| 2 | 276 (9.0%) |
| 3 | 228 (7.5%) |
| 4 | 89 (2.9%) |
| 5 | 86 (2.8%) |
| 6 | 43 (1.4%) |
| 7 | 113 (3.7%) |
| Légumes (jours) |  |
| 0 | 2,479 (81%) |
| 1 | 228 (7.5%) |
| 2 | 113 (3.7%) |
| 3 | 79 (2.6%) |
| 4 | 27 (0.9%) |
| 5 | 38 (1.2%) |
| 6 | 15 (0.5%) |
| 7 | 72 (2.4%) |
| Fruits (jours) |  |
| 0 | 2,179 (71%) |
| 1 | 397 (13%) |
| 2 | 231 (7.6%) |
| 3 | 124 (4.1%) |
| 4 | 34 (1.1%) |
| 5 | 31 (1.0%) |
| 6 | 12 (0.4%) |
| 7 | 40 (1.3%) |
| Viande/poisson/œufs (jours) |  |
| 0 | 2,099 (69%) |
| 1 | 244 (8.0%) |
| 2 | 243 (8.0%) |
| 3 | 175 (5.8%) |
| 4 | 91 (3.0%) |
| 5 | 64 (2.1%) |
| 6 | 15 (0.5%) |
| 7 | 112 (3.7%) |
| Produits laitiers (jours) |  |
| 0 | 2,775 (91%) |
| 1 | 56 (1.8%) |
| 2 | 56 (1.8%) |
| 3 | 47 (1.5%) |
| 4 | 25 (0.8%) |
| 5 | 38 (1.2%) |
| 6 | 3 (<0.1%) |
| 7 | 42 (1.4%) |
| Sucre (jours) |  |
| 0 | 2,144 (70%) |
| 1 | 152 (5.0%) |
| 2 | 198 (6.5%) |
| 3 | 199 (6.5%) |
| 4 | 109 (3.6%) |
| 5 | 70 (2.3%) |
| 6 | 41 (1.3%) |
| 7 | 134 (4.4%) |
| Matières grasses (jours) |  |
| 0 | 2,280 (75%) |
| 1 | 194 (6.4%) |
| 2 | 161 (5.3%) |
| 3 | 129 (4.2%) |
| 4 | 73 (2.4%) |
| 5 | 66 (2.2%) |
| 6 | 33 (1.1%) |
| 7 | 106 (3.5%) |
| **1n (%)** | |

### Calculer le score de consommation alimentaire

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

1. Tableau illustrant les poids attribués

Tableau : Groupes alimentaires et poids FCS

| **Groupe alimentaire** | **Exemples** | **Poids FCS** |
| --- | --- | --- |
| Céréales, tubercules | Riz, pain, manioc, igname | 2.0 |
| Légumineuses | Haricots, lentilles, pois | 3.0 |
| Produits laitiers | Lait, yaourt, fromage | 4.0 |
| Viandes/Poissons/Œufs | Viande, poisson, œufs | 4.0 |
| Légumes | Feuilles, gombo, carottes | 1.0 |
| Fruits | Mangue, banane, orange | 1.0 |
| Graisses/Huiles | Huile, beurre, margarine | 0.5 |
| Sucre | Sucre, miel, confiture | 0.5 |
| Condiments | Sel, épices, thé, café | 0.0 |

### Categoriser le SCA selon les seuil 21/35 et 28/42

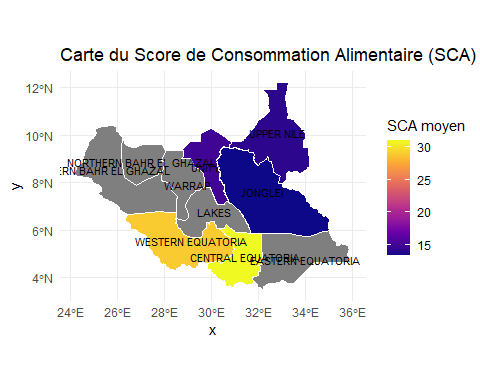
Score de consommation alimentaire

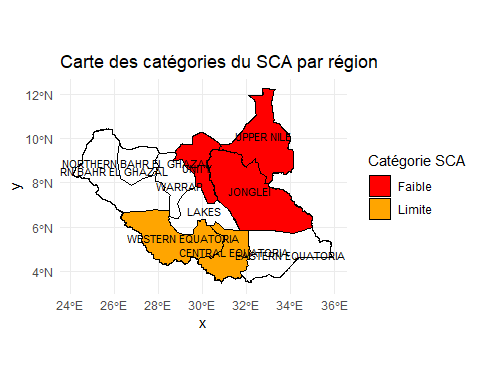
| **Characteristic** | **N = 3,058**1 |
| --- | --- |
| Days most hh members ate cereals, grains, roots and tubers Rice(Past 7days) | 17 (14, 0, 96) |
| Unknown | 23 |
| fcs\_cat\_28\_42 |  |
| Pauvre | 2,427 (79%) |
| Limite | 425 (14%) |
| Acceptable | 206 (6.7%) |
| fcs\_cat\_21\_35 |  |
| Pauvre | 2,085 (68%) |
| Limite | 603 (20%) |
| Acceptable | 370 (12%) |
| 1Mean (SD, Range); n (%) | |

## Faites une répresentation spatiale (région et département) du SCA et de ses différentes catégorisations

Toutes les noms de régions sont identiques.

[1] NA





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

L’indice réduit des stratégies de survie (rCSI) est un indicateur qui mesure les comportements d’adaptation que les ménages adoptent lorsqu’ils n’ont pas accès à suffisamment de nourriture. Il est basé sur un ensemble de cinq stratégies de survie communes liées à la consommation alimentaire. Un score plus élevé indique une plus grande insécurité alimentaire.

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

Variables du rCSI

| **Characteristic** | **N = 3,058**1 |
| --- | --- |
| A member ate less preferred food and less expensive food [Past 30/7 days] | 2,493 (82%) |
| Unknown | 2 |
| Any member borrowed food/relied on help to get more food...[Past 30/7 days] | 1,831 (60%) |
| Unknown | 5 |
| Any member ate a smaller meal than you felt you needed [Past 30/7 days] | 2,600 (85%) |
| Unknown | 4 |
| Any member ate fewer meals in a day [Past 30/7 days] | 2,547 (83%) |
| Unknown | 1 |
| Adults ate less to have more food for children under 5? [Past 30/7 days] | 2,145 (70%) |
| Unknown | 7 |
| 1n (%) | |

## b) Faites des statistiques descriptives sur les Nouvelles variables

## c) Calucler l’indice réduit des stratégies de survie en utilisant les Nouvelles variables

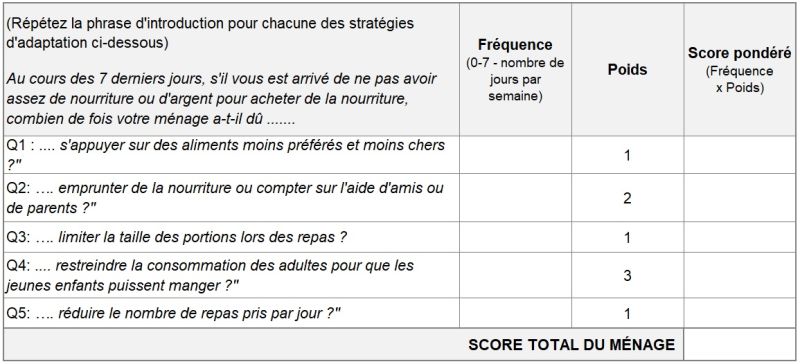
Tabulation de la variable

Indice réduit des stratégies de survie

| **Characteristic** | **N = 3,058**1 |
| --- | --- |
| Indice réduit des stratégies de survie | 23 (12, 0, 53) |
| Unknown | 15 |
| 1Mean (SD, Range) | |

## d) Faites un tableau illustrant le poids attribue a chaque variable pour le calcul du rCSI (la somme totale des poids doit etre egale a 21)

Cet tableau montre le poids de chaque variable.



## e) Faites une representation spatiale selon le milieu de residence du rCSI

Il n’y a pas de différence de nom dans les deux bases.

## e) Faites une representation spatiale selon le milieu de residence du rCSI

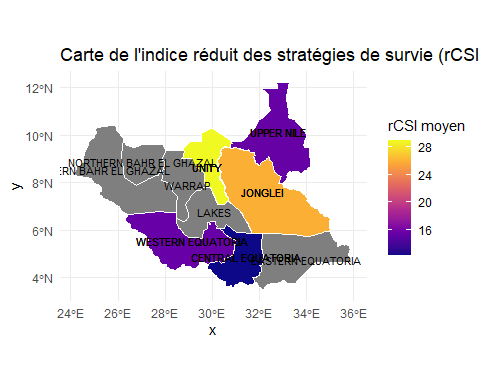


Table des matières

[I. Introduction 1](#_Toc197985273)

[II. I. Importation et Analyse de consistance des bases 2](#_Toc197985274)

[1. 1. Importation des jeux de données 2](#_Toc197985275)

[a. Base principale 2](#_Toc197985276)

[b. Base MAD 2](#_Toc197985277)

[III. II. Analyse de consistance des bases 2](#_Toc197985278)

[2. Detection de doublons 2](#_Toc197985279)

[3. Traitement de données manquantes 2](#_Toc197985280)

[IV. Analyse socio-economique des deplaces internes du Sudan en 2023 3](#_Toc197985281)

[4. 1) Pyramide des ages des individus 3](#_Toc197985282)

[5. 2) Caracteristiques du chef de menage 3](#_Toc197985283)

[6. 3) Crowding Index ou l’indice d’affluence 4](#_Toc197985284)

[7. a) Calculez le nombre d’individu par menage 4](#_Toc197985285)

[8. Faites des statistiques descriptives sur la variable renseigant le nombre de pieces du menage 5](#_Toc197985286)

[9. d) Calculez l’indice d’affluence ou crowding index 6](#_Toc197985287)

[10. e) Quelle est la proportion de menage dont le crowding index est: <1, 1<= crowding\_index<2, 2<= crowding\_index<3 et >=3 7](#_Toc197985288)

[11. Comparaison entre réfugiés et communautés d’accueil 7](#_Toc197985289)

[V. Analyse de la securite alimentaire des deplaces internes 8](#_Toc197985290)

[12. Score de consommation alimentaire (SCA) 8](#_Toc197985291)

[c. Faites une analyse descriptive des variables qui composent le SCA 9](#_Toc197985292)

[d. Calculer le score de consommation alimentaire 11](#_Toc197985293)

[e. Faites un tableau illustrant le poids attribue a chaque groupe alimentaire pour le calcul du SCA (la somme totale des poids doit etre egale a 16) 11](#_Toc197985294)

[f. Categoriser le SCA selon les seuil 21/35 et 28/42 12](#_Toc197985295)

[13. Faites une répresentation spatiale (région et département) du SCA et de ses différentes catégorisations 12](#_Toc197985296)

[VI. L’indice réduit des stratégies de survie (rCSI): 14](#_Toc197985297)

[14. a) Faites une analyse descriptive des variables qui composent le rCSI 14](#_Toc197985298)

[15. b) Faites des statistiques descriptives sur les Nouvelles variables 15](#_Toc197985299)

[16. c) Calucler l’indice réduit des stratégies de survie en utilisant les Nouvelles variables 15](#_Toc197985300)

[17. d) Faites un tableau illustrant le poids attribue a chaque variable pour le calcul du rCSI (la somme totale des poids doit etre egale a 21) 15](#_Toc197985301)

[18. e) Faites une representation spatiale selon le milieu de residence du rCSI 15](#_Toc197985302)

[19. e) Faites une representation spatiale selon le milieu de residence du rCSI 16](#_Toc197985303)

[VII. Table des matières 17](#_Toc197985304)