|  |
| --- |
| **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** |
| DIALLO Aissatou Sega | M. Aboubacar HEMA |
| *Élève Ingénieure Statisticienne Économiste* | *Data-scientist* |

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

# Sommaire

# Liste des figures

# Liste des tableaux

# Introduction

Les déplacements internes de population constituent aujourd’hui l’un des enjeux majeurs des crises humanitaires contemporaines, en particulier dans les pays confrontés à l’instabilité politique et aux conflits prolongés. Le Soudan du Sud, théâtre de tensions récurrentes, connaît une forte mobilité forcée de ses habitants, poussés à quitter leur lieu de résidence pour chercher refuge ailleurs. L’étude qui suit s’inscrit dans cette réalité préoccupante, à travers l’exploitation des données collectées lors de l’Enquête sur les déplacements forcés réalisée en 2023. Dans le cadre de notre cours « Projet statistique sur R », ce travail a pour objectif de mettre en œuvre des outils d’analyse statistique afin de dresser un portrait socio-économique et démographique des populations déplacées internes. À l’aide du logiciel R et dans une approche rigoureuse fondée sur la reproductibilité, il s’agira d’évaluer la qualité des bases de données, de caractériser les ménages, de mesurer les conditions d’habitation à travers l’indice d’affluence, et d’analyser la situation alimentaire par le biais du Score de Consommation Alimentaire (SCA) et de l’Indice réduit des Stratégies de Survie (rCSI). Au-delà des traitements quantitatifs, cette étude se propose également de comparer certaines dimensions entre les réfugiés et les communautés hôtes, en vue de mieux comprendre les dynamiques sociales en contexte de déplacement. Le présent rapport rend compte de ces analyses, illustrées par des graphiques, des tableaux de synthèse, et des commentaires interprétatifs appuyés sur les résultats statistiques obtenus sous R.

# Aperçu des données

Cette section propose une première exploration des données issues de l’enquête sur les déplacements forcés au Soudan du Sud en 2023. L’objectif est d’en présenter la structure générale et d’avoir un aperçu des principales variables disponibles. Pour ce faire, des fonctions de base de R seront utilisées afin de visualiser la nature des variables, leur type. Cette étape, bien que préliminaire, est cruciale pour orienter les traitements à venir.

## Chargement et importation

Avant toute analyse, il est nécessaire de charger les bibliothèques R utiles au traitement et à la visualisation des données. Ces packages offrent des fonctions pratiques pour la manipulation, l’exploration et la représentation graphique des données. Une fois les packages activés, les bases de données collectées lors de l’enquête sont importées dans l’environnement de travail.

#### Chargement des packages nécessaires

#### Chargement des bases de données

Dans cette étude, deux bases principales sont mobilisées : la base principale, contenant des informations au niveau des ménages (logement, conditions de vie, sécurité alimentaire, statut de réfugié, etc.), et la base individus, qui renseigne les caractéristiques socio-démographiques de chaque membre du ménage (âge, sexe, situation matrimoniale, etc.).

## Structure des données

### Dimension des bases de données

[1] 22092 144

La base Individus contient 22 092 personnes déplacées recensées au cours de l’enquête, chacune décrite par 144 variables. Ces variables peuvent inclure des informations sociodémographiques (âge, sexe, lien de parenté, niveau d’éducation), des expériences de déplacement, ou encore des besoins spécifiques liés à la situation humanitaire.

[1] 3058 1312

La base Principale contient 3 058 unités principales d’enquête, étant des ménages déplacés, des abris ou des familles, chacun décrit par 1 312 variables.

# I.Analyse consistence des bases

L’analyse de consistence des données constitue une étape importante dans tout projet statistique, visant à garantir la qualité, l’intégrité et la fiabilité des données avant d’engager toute analyse approfondie. Dans le contexte de l’Enquête sur les déplacements forcés au Sud-Soudan en 2023, cette analyse permet de détecter d’éventuelles anomalies, incohérences ou erreurs dans les bases de données, ce qui est essentiel pour assurer des résultats d’analyse corrects et pertinents. L’objectif de cette section est d’examiner la consistence des bases de données disponibles en vérifiant les valeurs manquantes, les doublons ou encore les incohérences logiques entre les variables.

## Standardisation des noms des variables

Avant toute manipulation des données, il est essentiel de s’assurer que les noms des variables sont standardisés, facilitant ainsi leur utilisation dans les analyses ultérieures.

Avant toute transformation, nous avons commencé par examiner les noms de variables des deux bases de données, Individus et Principale, afin de repérer d’éventuelles incohérences ou irrégularités (espaces, majuscules, caractères spéciaux, etc.). Cette étape préliminaire permet d’avoir une vision claire de l’état initial des noms de colonnes et de justifier l’utilité d’un nettoyage.

La fonction de nettoyage a ensuite été appliquée aux deux bases de données initiales. Les résultats ont été stockés dans deux nouveaux objets : individus\_clean et principale\_clean. Ces versions contiennent les mêmes données que les bases originales, mais avec des noms de colonnes nettoyés et standardisés.

Après nettoyage, les nouveaux noms de colonnes ont été affichés pour s’assurer que la transformation a bien été effectuée. Cette vérification visuelle permet de confirmer que les noms sont désormais harmonisés, lisibles et exploitables dans les étapes ultérieures d’analyse.

Nombre de noms modifiés dans la base Individus: 135

Nombre de noms modifiés dans la base Principale: 1297

Enfin, une vérification automatique a été réalisée pour comparer les noms de colonnes avant et après nettoyage. Le script calcule le nombre de noms modifiés dans chaque base. Cela permet de mesurer l’ampleur du nettoyage et de repérer s’il y a eu beaucoup de changements à apporter, ce qui peut révéler un besoin important de standardisation dans les données brutes.

## Gestion des doublons

La présence de doublons peut affecter significativement la qualité de l’analyse. Nous allons identifier et traiter les doublons au niveau des identifiants uniques et des observations complètes.

La base individu contient 0 doublons.

La base ménages contient 0 doublons.

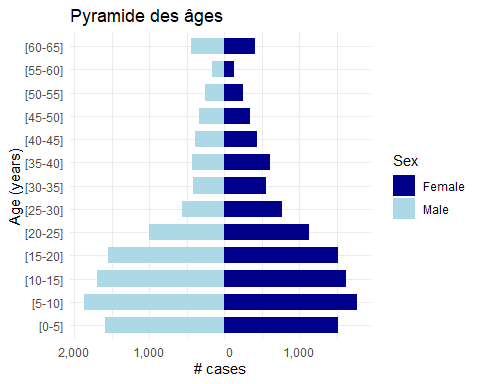
# II. Analyse socio-economique des deplaces internes du Sudan en 2023

Cette section vise à dresser un portrait détaillé des conditions socio-économiques des populations déplacées internes au Soudan en 2023, à partir des données recueillies lors de l’enquête. L’objectif est de mieux comprendre les caractéristiques démographiques, les conditions de logement, ainsi que la situation alimentaire des ménages affectés. À travers une série d’analyses descriptives et comparatives, cette étude met en lumière les différences entre les ménages déplacés (réfugiés) et les membres des communautés hôtes, notamment en termes de structure par âge, de statut matrimonial, de densité d’occupation des logements et de sécurité alimentaire.

## Pyramide des ages des individus

La pyramide des âges est un outil fondamental pour analyser la structure démographique d’une population. Elle permet de visualiser la répartition des individus par groupe d’âge et par sexe, mettant en évidence la jeunesse ou le vieillissement de la population ainsi que d’éventuels déséquilibres entre hommes et femmes. Dans le cadre de cette étude, la pyramide des âges des déplacés internes au Soudan en 2023 permet d’appréhender les dynamiques démographiques spécifiques à cette population, souvent marquée par des déplacements forcés, la désintégration familiale ou des risques sanitaires accrus.

La pyramide des âges issue des données de l’enquête sur les déplacements forcés au Soudan du Sud en 2023 révèle une population très jeune, avec une base large représentant les enfants et adolescents, et un sommet étroit traduisant une faible proportion de personnes âgées. Cette structure démographique est caractéristique des contextes de crise humanitaire, où les taux de natalité restent élevés mais où l’espérance de vie est réduite en raison des conditions de vie précaires, des conflits armés et d’un accès limité aux services de santé. La répartition par sexe est globalement équilibrée, bien que certaines tranches d’âge, notamment les 20-25 ans et les plus de 45 ans, montrent un excédent de femmes, ce qui peut résulter d’une mortalité plus élevée chez les hommes ou de déplacements différenciés selon le sexe. Les irrégularités visibles dans certaines tranches d’âge suggèrent l’impact direct des déplacements forcés, de la violence ou de la migration sélective sur la structure démographique. Cette pyramide met ainsi en évidence la vulnérabilité d’une population majoritairement composée d’enfants et de jeunes, avec des implications importantes pour la planification humanitaire et les politiques de protection.



**Figure** **:** Pyramide des ages des individus

## Caractéristiques des ménages

Cette partie permet de comparer certaines caractéristiques sociodémographiques des chefs de ménage issus de deux groupes distincts : les réfugiés et les membres des communautés hôtes du Nord. En s’appuyant sur la base de données de l’enquête sur les déplacements forcés menée au Sud-Soudan en 2023, seront extraits les individus concernés puis un résumé de statistiques clés sera effectués.

D’abord , on procède à une étape de filtrage et de préparation des données. On extrait uniquement les individus identifiés comme chefs de ménage grâce à la variable hh\_03. Puis, une nouvelle variable appelée groupe est créée à partir de la variable intro\_07\_1. Celle-ci permet de distinguer deux sous-populations : les réfugiés (valeur 1) et les membres de la communauté hôte du Nord (valeur 3). Les observations appartenant à d’autres groupes ou dont la modalité est manquante sont exclues.

On sélectionne ensuite quatre variables clés (groupe, âge, sexe, situation matrimoniale). À partir de ces variables, est construit un tableau statistique comparatif entre les deux groupes, présentant des moyennes pour les variables numériques et des fréquences pour les variables catégorielles, tout en ajoutant des tests de significativité.

Table 1: Caractéristiques des chefs de ménage

|  | *Group* | |  |
| --- | --- | --- | --- |
| *Caractéristique du Chef de ménage* | **Refugees** N = 2,061*1* | **Host community North** N = 988*1* | **p-value** |
| **Age du chef de ménage** | 42 (17, 61) | 42 (19, 61) |  |
| **Sexe du chef de ménage** |  |  |  |
| Male | 1,117 (54.2%) | 623 (63.1%) |  |
| Female | 942 (45.8%) | 365 (36.9%) |  |
| **Situation matrimoniale** |  |  |  |
| monogamous/married | 1,290 (62.7%) | 713 (72.2%) |  |
| polygamous/married | 282 (13.7%) | 186 (18.8%) |  |
| non-formal union | 12 (0.6%) | 3 (0.3%) |  |
| separated | 112 (5.4%) | 15 (1.5%) |  |
| divorced | 68 (3.3%) | 9 (0.9%) |  |
| widow or widower | 152 (7.4%) | 51 (5.2%) |  |
| never married | 143 (6.9%) | 11 (1.1%) |  |
| *1*Mean (Min, Max); n (%) | | | |
| Source : Enquête sur les déplacements forcés au Sud-Soudan en 2023 | | | |

Le tableau compare les caractéristiques des chefs de ménage parmi les réfugiés (N = 2 061) et la communauté hôte du nord (N = 988), selon l’enquête sur les déplacements forcés au Soudan du Sud (2023). L’âge médian est similaire dans les deux groupes (42 ans). Les réfugiés comptent proportionnellement plus de femmes chefs de ménage (45,8 % contre 36,9 %), ce qui peut indiquer une plus grande vulnérabilité. Concernant la situation matrimoniale, les chefs de ménage réfugiés sont moins souvent mariés (monogamie ou polygamie) et plus souvent séparés, divorcés, veufs ou jamais mariés que ceux de la communauté hôte. Ces différences suggèrent un impact social marqué des déplacements forcés, bien que l’absence de valeurs p empêche de juger de leur significativité statistique.

## Crowding Index ou l’indice d’affluence

L’indice d’affluence est le nombre de membres du menage divise par le nombre de pièces (à l’exclusion de la cuisine et des couloirs).

### Calcul du nombre d’individu par ménage

Nous avons créé une variable qui permet de retourner le nombre d’individus par ménage.

### Statistiques descriptives sur cette variable

Nous allons fournir des statistiques descriptives de la variable calculée précédemment.

| **Characteristic** | **N = 3,058**1 |
| --- | --- |
| nbre\_personnes | 7.1 (3.1) | Médiane: 7.0 | Min: 1.0 | Max: 17.0 |
| 1Mean (SD) | Médiane: Median | Min: Min | Max: Max | |
| Source: Calculs de l'auteur basés sur les données de l'enquête. | |

### Statistiques descriptives sur la variable renseignant le nombre de pieces du menage

| **Variable** | **Statistiques***1* |
| --- | --- |
| **Nombre de pièces du ménage** | 2.4 (1.3), médiane: 2, [1, 21.0] |
| Unknown | 6 |
| *1*Mean (SD), médiane: Median, [Min, Max] | |
| Source : Enquête sur les déplacements forcés au Sud-Soudan en 2023 | |

### Calcul de l’indice d’affluence ou crowding index

Dans cette partie , nous allons calculer l’indice d’affluence définie plus haut.

### Proportion de menage dont le crowding index est: <1, 1<= crowding\_index<2, 2<= crowding\_index<3 et >=3

| **Catégorie** | **Effectif et proportion***1* |
| --- | --- |
| **Catégorie de surpeuplement** |  |
| <1 | 68 (2.2%) |
| >=3 | 1,771 (58.0%) |
| 1-2 | 476 (15.6%) |
| 2-3 | 737 (24.1%) |
| *1*n (%) | |
| Source: Calculs de l'auteur basés sur les données de l'enquête. | |

# III. Analyse de la securite alimentaire des deplaces internes

La sécurité alimentaire constitue un enjeu central dans les situations de déplacement forcé, où l’accès régulier à une alimentation suffisante, diversifiée et nutritive est souvent compromis. Pour les populations déplacées internes, les perturbations liées aux conflits, à la perte de moyens de subsistance ou à l’absence d’infrastructures peuvent gravement affecter leur sécurité alimentaire. Cette section vise à évaluer la situation alimentaire des ménages déplacés au Soudan en 2023.

## Score de consommation alimentaire (SCA)

Le Score de Consommation Alimentaire (SCA) est un indicateur composite largement utilisé pour évaluer la sécurité alimentaire des ménages. Il repose sur trois dimensions essentielles : la diversité alimentaire, la fréquence de consommation des groupes d’aliments, et leur valeur nutritionnelle relative. Calculé sur une période de sept jours, il permet de mesurer dans quelle mesure les besoins alimentaires de base sont satisfaits.

### Analyse descriptive des variables qui composent le SCA

### Calcul du score de consommation alimentaire

Le calcul du Score de Consommation Alimentaire (SCA) repose sur la combinaison de plusieurs variables indiquant le nombre de jours, sur les sept derniers, durant lesquels les membres du ménage ont consommé différents groupes d’aliments. Chaque groupe alimentaire est pondéré en fonction de sa valeur nutritionnelle. Le score final est obtenu en multipliant la fréquence de consommation de chaque groupe par son poids correspondant, puis en faisant la somme de ces produits. Cette méthode permet d’obtenir un indicateur synthétique, reflétant à la fois la diversité, la fréquence et la qualité nutritionnelle de l’alimentation des ménages.

# Conclusion