Test Logiciel R

Aissatou Sega DIALLO

2025-02-25

# Test

Ce test se divise en trois parties : *-Gestion et nettoyage des bases de données*, *-Empilement et fusion des données,* *-Statistiques descriptives avec gtsummary.* Les bases de données utilisées dans ce test proviennent d’enquêtes menées auprès des mères afin d’analyser la composition nutritionnelle de leur alimentation ainsi que celle de leurs enfants au cours des 24 dernières heures. Les données ont été recueillies à deux moments distincts : au début de l’étude (Baseline) et à la fin de l’étude (Endline).

# Installation et chargement des packages

Cette section assure la disponibilité des bibliothèques nécessaires.

Packages <- c("readr", "haven", "utils", "dplyr","labelled", "gtsummary","skimr")  
  
for (x in Packages) {  
 if (!requireNamespace(x, quietly = TRUE)) {  
 install.packages(x)  
 }  
 library(x, character.only = TRUE)  
}

## Warning: le package 'haven' a été compilé avec la version R 4.3.3

## Warning: le package 'dplyr' a été compilé avec la version R 4.3.3

## Warning: le package 'gtsummary' a été compilé avec la version R 4.3.3

## Warning: le package 'skimr' a été compilé avec la version R 4.3.3

#Importation des bases de données

**Chargement des bases de données**

#Chargement des bases sur l'alimentation de la mère  
mother\_baseline<- haven::read\_dta("../Données/food\_comp\_mother\_baseline.dta")  
mother\_endline<- haven::read\_dta("../Données/food\_comp\_mother\_endline.dta")  
  
#Chargement des bases sur l'alimentation de l'enfant  
child\_baseline<- haven::read\_dta("../Données/food\_comp\_child\_baseline.dta")  
child\_endline<- haven::read\_dta("../Données/food\_comp\_child\_endline.dta")  
  
#Chargement de la base ménage   
Base\_men<- haven::read\_dta("../Données/base\_menage.dta")

**Aperçu des jeux de données**

# Mettre les bases dans une liste  
bases\_list\_init <- list(Base\_men, mother\_baseline, mother\_endline, child\_baseline, child\_endline)  
  
# Appliquez str() à chaque base avec une boucle for  
for (base in bases\_list\_init) {  
 utils::str(base)  
}

## tibble [1,065 × 21] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:1065] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:1065] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:1065] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:1065] "4948484848535052" "4948484848535053" "4948484848535055" "4948484848535056" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ hhsize : num [1:1065] 4 8 11 9 16 6 31 8 23 5 ...  
## ..- attr(\*, "label")= chr "Household size"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ poly : dbl+lbl [1:1065] 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0,...  
## ..@ label : chr "Polygamous household?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ hh\_primary : dbl+lbl [1:1065] 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Household head completed primary education"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s1\_q2 : dbl+lbl [1:1065] 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Male head of household"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Female" "Male"  
## $ s1\_q4a : num [1:1065] 37 52 67 38 75 41 56 52 63 26 ...  
## ..- attr(\*, "label")= chr "Age head of household"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ s2\_q1 : dbl+lbl [1:1065] 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Is head of household literate in local language"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s2\_q2 : dbl+lbl [1:1065] 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...  
## ..@ label : chr "Is head of household literate in French?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s2\_q4 : dbl+lbl [1:1065] 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0,...  
## ..@ label : chr "Has head of household been to school (formal or informal)?"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ s29\_q1 : dbl+lbl [1:1065] 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0,...  
## ..@ label : chr "Est-ce qu'un membre de votre ménage a pris un prêt ou fait un emprunt en argent"  
## ..@ format.stata: chr "%8.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Non" "Oui"  
## $ demgrp1 : num [1:1065] 1 0 0 0 1 0 5 1 1 1 ...  
## ..- attr(\*, "label")= chr "Number of children 0-36 months"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp2 : num [1:1065] 1 2 2 1 1 1 5 0 2 1 ...  
## ..- attr(\*, "label")= chr "Number of children 36-72 months"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp3 : num [1:1065] 0 3 3 3 7 2 11 4 9 1 ...  
## ..- attr(\*, "label")= chr "Number of adults 6-14 years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp4 : num [1:1065] 1 2 5 5 6 3 10 3 11 2 ...  
## ..- attr(\*, "label")= chr "Number of adults 14-65 years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ demgrp5 : num [1:1065] 0 1 1 0 1 0 0 0 0 0 ...  
## ..- attr(\*, "label")= chr "Number of elders 65+ years"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ dependencyratio: num [1:1065] 2 3 1.2 0.8 1.67 ...  
## ..- attr(\*, "label")= chr "Dependency ratio"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ hfias\_score : num [1:1065] 18 21 0 4 1 3 12 0 14 0 ...  
## ..- attr(\*, "label")= chr "HFIAS Score (0-27)"  
## ..- attr(\*, "format.stata")= chr "%9.0g"  
## $ T1 : dbl+lbl [1:1065] 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Treatment: 1st level comparison"  
## ..@ format.stata: chr "%12.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "Controle" "Intervention"  
## - attr(\*, "label")= chr "One row per household"  
## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ V1 : num [1:4256] 680 NA 634 256 563 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 23.31 NA 21.62 8.25 11.7 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 5.3 NA 4.87 8.06 6.49 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 62.7 NA 57.1 22 116.3 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 10.591 NA 9.897 0.912 2.716 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ V9 : num [1:4256] 4.507 NA 4.19 0.456 3.382 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.3058 NA 0.2835 0.0456 0.2248 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00869 NA 0.0078 0 0.00823 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 0.0441 NA 0.0396 0 0.0002 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 2217 1021 3038 716 618 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 48.7 34.8 92 19.1 16.5 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 89.6 5.94 183.6 4.73 1.63 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 818.8 72.1 1781.3 131.6 10.8 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 22.659 16.558 11.345 8.877 0.105 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 9.5691 6.4618 2.2278 3.5867 0.0334 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.8748 0.4327 0.314 0.5277 0.0059 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00948 0.02232 0 0.01281 0.01016 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 11.547 0.183 26.738 0.133 0 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 355 224 334 494 235 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 12.74 8.01 12.72 16.65 4.45 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 3.14 1.96 3.42 8.84 4.76 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 40.1 24.8 47 60.1 93.3 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 5.39 3.42 4.9 6.83 1.54 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 2.41 1.52 2.35 3.13 1.63 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.167 0.105 0.168 0.207 0.116 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00616 0.00377 0.00785 0.00959 0.00688 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 0.031282 0.019173 0.03986 0.048689 0.000167 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## tibble [4,256 × 17] (S3: tbl\_df/tbl/data.frame)  
## $ regionid : num [1:4256] 2 2 2 2 2 2 2 2 2 2 ...  
## ..- attr(\*, "label")= chr "Region ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ communeid : num [1:4256] 25 25 25 25 25 25 25 25 25 25 ...  
## ..- attr(\*, "label")= chr "Commune ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ villageid : num [1:4256] 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 ...  
## ..- attr(\*, "label")= chr "Village ID"  
## ..- attr(\*, "format.stata")= chr "%8.0g"  
## $ hhid : chr [1:4256] "4948484848535052" "4948484848535052" "4948484848535052" "4948484848535052" ...  
## ..- attr(\*, "label")= chr "Household ID"  
## ..- attr(\*, "format.stata")= chr "%45s"  
## $ round : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Survey round : Baseline, Endline"  
## ..@ format.stata: chr "%10.0g"  
## ..@ labels : Named num [1:2] 1 2  
## .. ..- attr(\*, "names")= chr [1:2] "Baseline" "Endline"  
## $ s1\_q0 : dbl+lbl [1:4256] 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4, 1, 2,...  
## ..@ label : chr "eating occasion"  
## ..@ format.stata: chr "%27.0g"  
## ..@ labels : Named num [1:4] 1 2 3 4  
## .. ..- attr(\*, "names")= chr [1:4] "Breakfast" "Lunch" "Dinner" "Snacks"  
## $ s1\_q1 : dbl+lbl [1:4256] 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## ..@ label : chr "Meal consumed? Y/N"  
## ..@ format.stata: chr "%9.0g"  
## ..@ labels : Named num [1:2] 0 1  
## .. ..- attr(\*, "names")= chr [1:2] "No" "Yes"  
## $ s1\_q2 : dbl+lbl [1:4256] 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,...  
## ..@ label : chr "Who consummed?"  
## ..@ format.stata: chr "%16.0g"  
## ..@ labels : Named num [1:3] 1 2 3  
## .. ..- attr(\*, "names")= chr [1:3] "mother" "child" "mother and child"  
## $ energ\_kcal : num [1:4256] 1193 535 1701 531 188 ...  
## ..- attr(\*, "label")= chr "Consommation en ernergie (kcal)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ protein\_g : num [1:4256] 24.94 18.38 58.84 13.89 5.04 ...  
## ..- attr(\*, "label")= chr "Consommation en proteine (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ lipid\_tot\_g: num [1:4256] 51.915 3.205 90.268 3.917 0.496 ...  
## ..- attr(\*, "label")= chr "Consommation en lipide (g)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ calcium\_mg : num [1:4256] 479.73 41.32 1122.63 74.44 3.29 ...  
## ..- attr(\*, "label")= chr "Consommation en calcium (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ iron\_mg : num [1:4256] 11.5857 8.709 7.2277 6.4328 0.0321 ...  
## ..- attr(\*, "label")= chr "Consommation en fer (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ zinc\_mg : num [1:4256] 4.9181 3.3562 1.4184 2.7147 0.0102 ...  
## ..- attr(\*, "label")= chr "Consommation en zinc (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b6\_mg : num [1:4256] 0.4678 0.2251 0.2028 0.3523 0.0018 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B6 (mg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_b12\_mcg: num [1:4256] 0.00396 0.01364 0 0.00736 0.00309 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine B12 (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"  
## $ vit\_c\_mg : num [1:4256] 6.8299 0.1201 16.904 0.0764 0 ...  
## ..- attr(\*, "label")= chr "Consommation en vitamine C (mcg)"  
## ..- attr(\*, "format.stata")= chr "%10.0g"

#**Partie 1 : Gestion et nettoyage des bases de données** Dans cette première partie, nous allons effectuer une série d’opérations pour garantir la qualité des bases de données en les nettoyant et en les harmonisant. Cela comprend la vérification et la suppression des doublons, l’harmonisation des noms de variables, la gestion des valeurs manquantes, ainsi que le calcul de certaines statistiques descriptives.

##1.Vérification et suppression des données

#Suppression des doublons   
for (i in 1:length(bases\_list\_init)) {  
 bases\_list\_init[[i]] <- bases\_list\_init[[i]] %>% distinct()  
}

##2.Harmonisation des variables

# Vérification des noms de variables pour chaque base  
for (base in bases\_list\_init) {  
 print(colnames(base))  
}

## [1] "regionid" "communeid" "villageid" "hhid"   
## [5] "hhsize" "poly" "hh\_primary" "s1\_q2"   
## [9] "s1\_q4a" "s2\_q1" "s2\_q2" "s2\_q4"   
## [13] "s29\_q1" "demgrp1" "demgrp2" "demgrp3"   
## [17] "demgrp4" "demgrp5" "dependencyratio" "hfias\_score"   
## [21] "T1"   
## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "V1" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "V9" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"   
## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"   
## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"   
## [1] "regionid" "communeid" "villageid" "hhid" "round"   
## [6] "s1\_q0" "s1\_q1" "s1\_q2" "energ\_kcal" "protein\_g"   
## [11] "lipid\_tot\_g" "calcium\_mg" "iron\_mg" "zinc\_mg" "vit\_b6\_mg"   
## [16] "vit\_b12\_mcg" "vit\_c\_mg"

# Liste des bases de données Baseline et Endline  
baseline\_list <- list(mother\_baseline, child\_baseline)   
endline\_list <- list(mother\_endline, child\_endline)   
  
# Liste des noms de variables dans Baseline et Endline  
baseline\_names <- lapply(baseline\_list, colnames)  
endline\_names <- lapply(endline\_list, colnames)  
  
# Comparaison des noms de variables entre Baseline et Endline  
for (i in 1:length(baseline\_list)) {  
 baseline\_vars <- baseline\_names[[i]]  
 endline\_vars <- endline\_names[[i]]  
   
 # Si les noms des variables ne sont pas identiques, les renommer  
 if (!identical(baseline\_vars, endline\_vars)) {  
 # Créez une liste de correspondance des noms de variables à renommer   
 rename\_mapping <- setNames(baseline\_vars, endline\_vars)  
   
 # Renommer les variables dans la base Endline  
 endline\_list[[i]] <- endline\_list[[i]] %>%  
 rename\_with(~ rename\_mapping[.], everything()) # Applique le renommage  
 }  
}

##3.Gestion des valeurs manquantes

# Vérification des valeurs manquantes avec skim()  
for (base in bases\_list\_init) {  
 skim(base)  
}  
  
# Imputation des valeurs manquantes par la moyenne pour chaque base  
for (i in 1:length(bases\_list\_init)) {  
 bases\_list\_init[[i]] <- bases\_list\_init[[i]] %>%   
 mutate(across(everything(), ~ifelse(is.na(.), mean(., na.rm = TRUE), .)))  
}

##4.Consommation d’énergie moyenne à chaque repas pour l’ensemble des mères lors de l’enquête de base

# Calcul de la moyenne de consommation d'énergie  
moy\_energie\_mother <- mother\_baseline %>%  
 group\_by(s1\_q0) %>%  
 summarise(moy\_energie = mean(V1, na.rm = TRUE))  
  
print(moy\_energie\_mother)

## # A tibble: 4 × 2  
## s1\_q0 moy\_energie  
## <dbl+lbl> <dbl>  
## 1 1 [Breakfast] 637.  
## 2 2 [Lunch] 771.  
## 3 3 [Dinner] 793.  
## 4 4 [Snacks] 622.

##5.Renommage des bases corrigées

data\_names <- c("base\_menage\_final.dta", "mother\_baseline\_v1.dta", "mother\_endline\_v1.dta", "child\_baseline\_v1.dta", "child\_endline\_v1.dta")  
  
# Créer de nouvelles variables avec les noms souhaités  
for (i in 1:length(bases\_list\_init)) {  
 # Créer une nouvelle variable avec le nom correspondant  
 assign(data\_names[i], bases\_list\_init[[i]])   
 # Enregistrer la base sous son nouveau nom  
 write\_dta(get(data\_names[i]), data\_names[i])  
}

#**Partie 2 : Empilement et Fusion des données** Cette partie,après avoir vérifié et corrigé les possibles incohérences dans les bases de données , permettra de faire l’empilement et la fusion des jeux de données.

##Utilisation des bases Baseline

**Empilation des bases Baseline**

#Empilation des bases mères et enfants  
mother\_baseline\_v1.dta <- mother\_baseline\_v1.dta %>%  
 rename(energ\_kcal = V1, # Renomme 'old\_name\_1' en 'new\_name\_1'  
 zinc\_mg = V9) # Renomme 'old\_name\_2' en 'new\_name\_2'  
  
baseline <- rbind(mother\_baseline\_v1.dta ,child\_baseline\_v1.dta)

**Renommage des variables**

#Renommer les variables de consommation  
# rename(energ\_kcal = V1) # Renomme la variable V1 en energ\_kcal  
#rename(zinc\_mg = V9) # Renomme la variable V9 en zinc\_mg  
baseline <- baseline %>%   
 rename\_with(~paste0(., "\_b"),   
 .cols = energ\_kcal:vit\_c\_mg)

**Création d’une base de données synthétisante**

baseline\_summary <- merge(  
 baseline %>%   
 group\_by(hhid) %>%  
 summarise(across(s1\_q2:vit\_c\_mg\_b, sum, na.rm = TRUE), .groups = "drop"), # Ajout de .groups = "drop"  
   
 baseline %>%   
 select(hhid) %>%   
 distinct(),  
   
 by = c("hhid"),  
 all = TRUE  
)

## Warning: There was 1 warning in `summarise()`.  
## ℹ In argument: `across(s1\_q2:vit\_c\_mg\_b, sum, na.rm = TRUE)`.  
## ℹ In group 1: `hhid = "4948484848535052"`.  
## Caused by warning:  
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.  
## Supply arguments directly to `.fns` through an anonymous function instead.  
##   
## # Previously  
## across(a:b, mean, na.rm = TRUE)  
##   
## # Now  
## across(a:b, \(x) mean(x, na.rm = TRUE))

**Sauvegarder les données dans une base finale**

baseline\_final <- baseline\_summary %>%   
 select(hhid, s1\_q2, energ\_kcal\_b, protein\_g\_b, lipid\_tot\_g\_b,   
 calcium\_mg\_b, iron\_mg\_b, zinc\_mg\_b)  
write\_dta(baseline\_final, "baseline\_final.dta")

##Utilisation des bases Endline **Empilation des bases et renommage**

#Empilation des bases endline  
  
  
endline <- rbind(mother\_endline\_v1.dta ,child\_endline\_v1.dta)  
  
#Renommage des variables  
endline <- endline %>%   
 rename\_with(~paste0(., "\_e"),   
 .cols = energ\_kcal:vit\_c\_mg)

**Création d’une base de données synthétisante**

endline\_summary <- endline %>%   
 group\_by(hhid) %>%   
 summarise(across(s1\_q2:vit\_c\_mg\_e, sum, na.rm = TRUE), .groups = "drop")  
  
colnames(endline\_summary)

## [1] "hhid" "s1\_q2" "energ\_kcal\_e" "protein\_g\_e"   
## [5] "lipid\_tot\_g\_e" "calcium\_mg\_e" "iron\_mg\_e" "zinc\_mg\_e"   
## [9] "vit\_b6\_mg\_e" "vit\_b12\_mcg\_e" "vit\_c\_mg\_e"

**Sauvegarder les données dans une base finale**

endline\_final <- endline\_summary %>%   
 select(hhid, s1\_q2, energ\_kcal\_e, protein\_g\_e, lipid\_tot\_g\_e,   
 calcium\_mg\_e, iron\_mg\_e, zinc\_mg\_e)  
write\_dta(endline\_final, "endline\_final.dta")

##Fusion des bases Endline et Baseline

#Fusion des bases Baseline ET Endline  
final\_data <- left\_join(baseline\_final, endline\_final, by = c("hhid", "s1\_q2"))  
write\_dta(final\_data, "final\_data.dta")

##Fusion avec les caractéristiques des ménages

base\_menage\_final\_selected <- base\_menage\_final.dta %>%  
 select(hhid, hhsize, regionid, communeid, villageid, poly, dependencyratio, hfias\_score, T1)  
base\_finale <- left\_join(final\_data, base\_menage\_final\_selected, by = "hhid")  
write\_dta(base\_finale, "base\_finale.dta")

#**Partie 3 : Statistiques descriptives avec le package gtsummary**

**Création d’un tableau récapitulatif de consommation**

# Création d'un tableau de statistiques descriptives  
table\_summary <- base\_finale %>%  
 select(energ\_kcal\_b, protein\_g\_b, lipid\_tot\_g\_b, calcium\_mg\_b, iron\_mg\_b, zinc\_mg\_b,  
 energ\_kcal\_e, protein\_g\_e, lipid\_tot\_g\_e, calcium\_mg\_e, iron\_mg\_e, zinc\_mg\_e) %>%  
 tbl\_summary(statistic = list(all\_continuous() ~ "{mean} ({sd})"),  
 missing = "no") %>%  
modify\_header(label="Présentation de la consommation moyenne selon la période")  
table\_summary

| Présentation de la consommation moyenne selon la période | **N = 1,064***1* |
| --- | --- |
| energ\_kcal\_b | 4,509 (1,867) |
| protein\_g\_b | 140 (103) |
| lipid\_tot\_g\_b | 78 (74) |
| calcium\_mg\_b | 778 (644) |
| iron\_mg\_b | 57 (41) |
| zinc\_mg\_b | 23 (12) |
| energ\_kcal\_e | 4,601 (2,172) |
| protein\_g\_e | 140 (75) |
| lipid\_tot\_g\_e | 90 (77) |
| calcium\_mg\_e | 864 (885) |
| iron\_mg\_e | 50 (34) |
| zinc\_mg\_e | 21 (10) |
| *1*Mean (SD) | |

**Croisements avec certaines informations du ménage**

#Consommation en nutriments selon le statut matrimonial (polygamie ou non)  
table\_matrimonial <- base\_finale %>%  
 select(poly, energ\_kcal\_b, protein\_g\_b, lipid\_tot\_g\_b) %>%  
 tbl\_summary(by = poly, statistic = all\_continuous() ~ "{mean} ({sd})") %>%  
 modify\_header(label="Consommation moyenne en début de période selon le régime matrimonial")  
table\_matrimonial

| Consommation moyenne en début de période selon le régime matrimonial | **0** N = 554*1* | **1** N = 510*1* |
| --- | --- | --- |
| energ\_kcal\_b | 4,500 (1,924) | 4,519 (1,805) |
| protein\_g\_b | 140 (115) | 140 (88) |
| lipid\_tot\_g\_b | 78 (79) | 79 (69) |
| *1*Mean (SD) | | |

#Consommation en nutriments selon la région  
table\_region <- base\_finale %>%  
 select(regionid, energ\_kcal\_b, protein\_g\_b, lipid\_tot\_g\_b) %>%  
 tbl\_summary(by = regionid, statistic = all\_continuous() ~ "{mean} ({sd})") %>%  
 modify\_header(label="Consommation moyenne en début de période selon le région")  
table\_region

| Consommation moyenne en début de période selon le région | **1** N = 188*1* | **2** N = 619*1* | **3** N = 257*1* |
| --- | --- | --- | --- |
| energ\_kcal\_b | 4,261 (2,052) | 4,530 (1,886) | 4,639 (1,657) |
| protein\_g\_b | 133 (114) | 142 (71) | 141 (151) |
| lipid\_tot\_g\_b | 73 (101) | 83 (73) | 70 (51) |
| *1*Mean (SD) | | | |

# Regroupement des tailles de ménage en classes  
base\_finale <- base\_finale %>%   
 mutate(taille\_menage\_classe = case\_when(  
 hhsize <= 2 ~ "Petite (1-2)",  
 hhsize <= 5 ~ "Moyenne (3-5)",  
 hhsize <= 8 ~ "Grande (6-8)",  
 TRUE ~ "Très grande (9+)")  
 )  
  
#Consommation en nutriments selon la taille du ménage  
table\_size <- base\_finale %>%  
 select(taille\_menage\_classe, energ\_kcal\_b, protein\_g\_b, lipid\_tot\_g\_b, calcium\_mg\_b) %>%  
 tbl\_summary(by = taille\_menage\_classe, statistic = all\_continuous() ~ "{mean} ({sd})") %>%  
 modify\_header(label="Consommation moyenne en début de période selon la taille du ménage")  
table\_size

| Consommation moyenne en début de période selon la taille du ménage | **Grande (6-8)** N = 333*1* | **Moyenne (3-5)** N = 260*1* | **Petite (1-2)** N = 3*1* | **Très grande (9+)** N = 468*1* |
| --- | --- | --- | --- | --- |
| energ\_kcal\_b | 4,441 (1,759) | 4,374 (1,896) | 3,569 (697) | 4,638 (1,924) |
| protein\_g\_b | 140 (133) | 137 (108) | 88 (18) | 143 (72) |
| lipid\_tot\_g\_b | 77 (68) | 77 (91) | 42 (9) | 81 (69) |
| calcium\_mg\_b | 743 (585) | 670 (515) | 395 (123) | 865 (734) |
| *1*Mean (SD) | | | | |