Untitled

KAFANDO

2025-02-25

# PARTIE 0 :

## **Installation et importation des packages**

Dans cette section, nous installons tous les packages qui servirons dans la suite.

packages <- c("readr","cardx","haven","utils","dplyr","gtsummary","gt","labelled")  
  
  
for (package in packages) {  
 if (!requireNamespace(package, quietly = TRUE)) { # Vérifie si le package n'est pas encore installé  
 install.packages(package)  
 }  
 library(package, character.only = TRUE) # nom du package en nom ou chaine de caractère ()  
}

## Importation des bases

##  
base\_menage <- haven::read\_dta("../Données/base\_menage.dta") # Importation de la base ménage  
  
  
food\_comp\_child\_baseline <- haven::read\_dta("../Données/food\_comp\_child\_baseline.dta") # Importation de la base food\_comp\_child\_baseline  
  
food\_comp\_child\_endline <- haven::read\_dta("../Données/food\_comp\_child\_endline.dta") # Importation de la base food\_comp\_child\_endline  
  
food\_comp\_mother\_baseline <- haven::read\_dta("../Données/food\_comp\_mother\_baseline.dta") # Importation de la base food\_comp\_mother\_baseline  
  
food\_comp\_mother\_endline <- haven::read\_dta("../Données/food\_comp\_mother\_endline.dta") # Importation de la base food\_comp\_mother\_endline

## Analyse des bases

**Structure de chaque base**

utils::str(base\_menage) # Structure de la base ménage

## 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"

utils::str(food\_comp\_child\_baseline) # Structure de la base food\_comp\_child\_baseline

## 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"

utils::str(food\_comp\_child\_endline) # Structure de la base food\_comp\_child\_endline

## 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"

utils::str(food\_comp\_mother\_baseline) # Structure de la base food\_comp\_mother\_baseline

## 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"

utils::str(food\_comp\_mother\_endline) # Structure de la base food\_comp\_mother\_endline

## 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"

**Noms des colonnes de chaque base**

base::colnames(base\_menage) # Nom des colonnes de la base ménage

## [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"

base::colnames(food\_comp\_child\_baseline) # Nom des colonne de la base food\_comp\_child\_baseline

## [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"

base::colnames(food\_comp\_child\_endline) # Nom des colonne de la base food\_comp\_child\_endline

## [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"

base::colnames(food\_comp\_mother\_baseline) # Nom des colonne de la base food\_comp\_mother\_baseline

## [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"

base::colnames(food\_comp\_mother\_endline) # Nom des colonne de la base food\_comp\_mother\_endline

## [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"

# Partie 1 : Gestion et nettoyage des bases de données

## 1. Vérifiez la présence de doublons dans les bases de données Baseline, Endline et ménage. Supprimez les doublons si nécessaire.

cat("Nombre d'observations dans base ménage avant suppression des doublons :", nrow(base\_menage), "\n")

## Nombre d'observations dans base ménage avant suppression des doublons : 1065

cat("Nombre d'observations dans food\_comp\_child\_baseline avant suppression des doublons :", nrow(food\_comp\_child\_baseline), "\n")

## Nombre d'observations dans food\_comp\_child\_baseline avant suppression des doublons : 4256

cat("Nombre d'observations dans food\_comp\_child\_endline avant suppression des doublons :", nrow(food\_comp\_child\_endline), "\n")

## Nombre d'observations dans food\_comp\_child\_endline avant suppression des doublons : 4256

cat("Nombre d'observations dans food\_comp\_mother\_baseline avant suppression des doublons :", nrow(food\_comp\_mother\_baseline), "\n")

## Nombre d'observations dans food\_comp\_mother\_baseline avant suppression des doublons : 4256

cat("Nombre d'observations dans food\_comp\_mother\_endline avant suppression des doublons :", nrow(food\_comp\_mother\_endline), "\n")

## Nombre d'observations dans food\_comp\_mother\_endline avant suppression des doublons : 4256

# Supprimer les doublons en se basant sur l'identifiant du ménage  
base\_menage = base\_menage %>% distinct(hhid, .keep\_all = TRUE)  
  
cat("Nombre d'observations dans base\_menage après suppression des doublons :", nrow(base\_menage), "\n")

## Nombre d'observations dans base\_menage après suppression des doublons : 1064

# Supprimer les doublons dans food\_comp\_child\_baseline  
food\_comp\_child\_baseline = dplyr::distinct(food\_comp\_child\_baseline)  
  
cat("Nombre d'observations dans food\_comp\_child\_baseline après suppression des doublons :", nrow(food\_comp\_child\_baseline), "\n")

## Nombre d'observations dans food\_comp\_child\_baseline après suppression des doublons : 4256

# Supprimer les doublons dans food\_comp\_child\_endline  
food\_comp\_child\_endline = dplyr::distinct(food\_comp\_child\_endline)  
  
cat("Nombre d'observations dans food\_comp\_child\_endline après suppression des doublons :", nrow(food\_comp\_child\_endline), "\n")

## Nombre d'observations dans food\_comp\_child\_endline après suppression des doublons : 4256

# Supprimer les doublons dans food\_comp\_mother\_baseline  
food\_comp\_mother\_baseline = dplyr::distinct(food\_comp\_mother\_baseline)  
  
cat("Nombre d'observations dans food\_comp\_mother\_baseline après suppression des doublons :", nrow(food\_comp\_mother\_baseline), "\n")

## Nombre d'observations dans food\_comp\_mother\_baseline après suppression des doublons : 4256

# Supprimer les doublons dans food\_comp\_mother\_endline  
food\_comp\_mother\_endline = dplyr::distinct(food\_comp\_mother\_endline)  
  
cat("Nombre d'observations dans food\_comp\_mother\_endline après suppression des doublons :", nrow(food\_comp\_mother\_endline), "\n")

## Nombre d'observations dans food\_comp\_mother\_endline après suppression des doublons : 4256

## 2. Assurez-vous que les noms des variables sont cohérents entre les bases de données Baseline et Endline (par exemple, les noms des nutriments doivent être identiques). Si ce n’est pas le cas, apportez des corrections.

base::colnames(food\_comp\_child\_baseline) # Nom des colonne de la base food\_comp\_child\_baseline

## [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"

base::colnames(food\_comp\_child\_endline) # Nom des colonne de la base food\_comp\_child\_endline

## [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"

base::colnames(food\_comp\_mother\_baseline) # Nom des colonne de la base food\_comp\_mother\_baseline

## [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"

base::colnames(food\_comp\_mother\_endline) # Nom des colonne de la base food\_comp\_mother\_endline

## [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"

**Ici, le nom des variables n’est pas pareil**

# Renommer les colonnes V9 et V1  
  
food\_comp\_mother\_baseline <- food\_comp\_mother\_baseline %>% rename(zinc\_mg = V9, energ\_kcal = V1)

## 3. Certaines variables peuvent contenir des données manquantes, cependant suivant une logique bien établie, il est possible de corriger ces données manquantes. Veuillez vérifier soigneusement les données et corriger les données manquantes de certaines variables si possibles.

# Calculer le nombre de valeurs manquantes par variable (colonne)  
print("Valeurs manquante base ménage")

## [1] "Valeurs manquante base ménage"

NA\_base\_men <- colSums(is.na(base\_menage))  
NA\_base\_men

## regionid communeid villageid hhid hhsize   
## 0 1 0 0 0   
## poly hh\_primary s1\_q2 s1\_q4a s2\_q1   
## 0 0 0 0 0   
## s2\_q2 s2\_q4 s29\_q1 demgrp1 demgrp2   
## 0 0 0 0 0   
## demgrp3 demgrp4 demgrp5 dependencyratio hfias\_score   
## 0 0 0 0 0   
## T1   
## 0

print("Valeurs manquante base food\_comp\_mother\_baseline")

## [1] "Valeurs manquante base food\_comp\_mother\_baseline"

NA\_mother\_base <- colSums(is.na(food\_comp\_mother\_baseline))  
NA\_mother\_base

## regionid communeid villageid hhid round s1\_q0   
## 4 8 0 0 0 0   
## s1\_q1 s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg   
## 0 0 673 673 673 673   
## iron\_mg zinc\_mg vit\_b6\_mg vit\_b12\_mcg vit\_c\_mg   
## 673 673 673 673 673

print("Valeurs manquante base food\_comp\_mother\_endline")

## [1] "Valeurs manquante base food\_comp\_mother\_endline"

NA\_mother\_end <- colSums(is.na(food\_comp\_mother\_endline))  
NA\_mother\_end

## regionid communeid villageid hhid round s1\_q0   
## 0 0 0 0 0 0   
## s1\_q1 s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg   
## 0 0 664 664 664 664   
## iron\_mg zinc\_mg vit\_b6\_mg vit\_b12\_mcg vit\_c\_mg   
## 664 664 664 664 664

print("Valeurs manquante base food\_comp\_child\_baseline")

## [1] "Valeurs manquante base food\_comp\_child\_baseline"

NA\_child\_base <- colSums(is.na(food\_comp\_child\_baseline))  
NA\_child\_base

## regionid communeid villageid hhid round s1\_q0   
## 0 0 0 0 0 0   
## s1\_q1 s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg   
## 0 0 386 386 386 386   
## iron\_mg zinc\_mg vit\_b6\_mg vit\_b12\_mcg vit\_c\_mg   
## 386 386 386 386 386

print("Valeurs manquante base food\_comp\_child\_endline")

## [1] "Valeurs manquante base food\_comp\_child\_endline"

NA\_child\_end <- colSums(is.na(food\_comp\_child\_endline))  
NA\_child\_end

## regionid communeid villageid hhid round s1\_q0   
## 0 0 0 0 0 0   
## s1\_q1 s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg   
## 0 0 532 532 532 532   
## iron\_mg zinc\_mg vit\_b6\_mg vit\_b12\_mcg vit\_c\_mg   
## 532 532 532 532 532

**Traitons les valeurs manquantes dans région et commune de la base food\_comp\_mother\_baseline**

# 1. Remplacer les valeurs manquantes dans 'commune' en se basant sur le village  
food\_comp\_mother\_baseline <- food\_comp\_mother\_baseline %>%  
 group\_by(villageid) %>% # Grouper par village  
 mutate(communeid = ifelse(is.na(communeid), first(communeid[!is.na(communeid)]), communeid)) %>% #les individus n'ayant pas de commune prennent la commune de celui du meme village que lui  
 ungroup() # Dégrouper  
  
# 2. Remplacer les valeurs manquantes dans 'region' en se basant sur la commune  
  
food\_comp\_mother\_baseline <- food\_comp\_mother\_baseline %>%  
 group\_by(communeid) %>% # grouper par commune  
 mutate(regionid = ifelse(is.na(regionid), first(regionid[!is.na(regionid)]), regionid)) %>% #les individus n'ayant pas de region prennent la région de celui du meme commune que lui  
 ungroup()# Dégrouper

**Vérification**

print("Valeurs manquante base food\_comp\_mother\_baseline")

## [1] "Valeurs manquante base food\_comp\_mother\_baseline"

NA\_mother\_base <- colSums(is.na(food\_comp\_mother\_baseline))  
NA\_mother\_base

## regionid communeid villageid hhid round s1\_q0   
## 0 0 0 0 0 0   
## s1\_q1 s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg   
## 0 0 673 673 673 673   
## iron\_mg zinc\_mg vit\_b6\_mg vit\_b12\_mcg vit\_c\_mg   
## 673 673 673 673 673

**Traitement fait avec succès**

## 4. Quelle est la consommation d’énergie moyenne à chaque repas pour l’ensemble des mères lors de l’enquête de base ?

conso\_energie\_moy\_base <- food\_comp\_mother\_baseline %>%   
 labelled::to\_factor() %>% # labellisation des modalités  
 group\_by(s1\_q0) %>% # grouper suivant les 4 repas  
 summarise(moyenne = mean(energ\_kcal,na.rm=TRUE)) %>% # faire la moyenne en ne considérant pas la NA  
 ungroup()  
conso\_energie\_moy\_base

## # A tibble: 4 × 2  
## s1\_q0 moyenne  
## <fct> <dbl>  
## 1 Breakfast 637.  
## 2 Lunch 771.  
## 3 Dinner 793.  
## 4 Snacks 622.

## 5. Après vérification et correction de chaque base de données :

- i) Sauvegarder les changements de la base food\_comp\_mother\_baseline.dta en mother\_baseline\_v1.dta

dossier\_sauvegarde <- "../Données/" # spécification du dossier de sauvegarde  
  
fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "mother\_baseline\_v1.dta") # Création du dossier complet  
  
haven::write\_dta(food\_comp\_mother\_baseline, fichier\_sauvegarde) # Sauvegarde

- ii) Sauvegarder les changements de la base food\_comp\_mother\_endline.dta en mother\_endline\_v1.dta

fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "mother\_endline\_v1.dta") # Création du dossier complet  
  
haven::write\_dta(food\_comp\_mother\_endline, fichier\_sauvegarde) # Sauvegarde

- iii) Sauvegarder les changements de la base food\_comp\_child\_baseline.dta en child\_baseline\_v1.dta

fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "child\_baseline\_v1.dta") # Création du dossier complet  
  
haven::write\_dta(food\_comp\_child\_baseline, fichier\_sauvegarde) # Sauvegarde

- iv) Sauvegarder les changements de la base food\_comp\_child\_endline.dta en child\_endline\_v1.dta

fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "child\_endline\_v1.dta") # Création du dossier complet  
  
haven::write\_dta(food\_comp\_child\_endline, fichier\_sauvegarde) # Sauvegarde

- v) Sauvegarder les changements de la base base\_menage.dta en base\_menage\_final.dta

fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "base\_menage\_final.dta") # Création du dossier complet  
  
haven::write\_dta(base\_menage, fichier\_sauvegarde) # Sauvegarde

# **Partie 2 : Empilement et Fusion des données**

*Maintenant que vous avez vérifié et corrigé les possibles incohérences dans les bases de données suivre les instructions suivantes :*

## 1) **Baseline**

**Importons les différentes bases**

##  
base\_menage\_final <- haven::read\_dta("../Données/base\_menage\_final.dta") # Importation de la base ménage finale  
  
child\_baseline\_v1 <- haven::read\_dta("../Données/child\_baseline\_v1.dta") # Importation de la base child\_baseline\_v1.dta  
  
child\_endline\_v1 <- haven::read\_dta("../Données/child\_endline\_v1.dta") # Importation de la base child\_endline\_v1.dta  
  
mother\_baseline\_v1 <- haven::read\_dta("../Données/mother\_baseline\_v1.dta") # Importation de la base mother\_baseline\_v1.dta  
  
mother\_endline\_v1.dta <- haven::read\_dta("../Données/mother\_endline\_v1.dta") # Importation de la base mother\_endline\_v1

* *i) Empilez les bases de données des mères (mother\_baseline\_v1.dta) et des enfants (child\_baseline\_v1.dta) de sorte que pour chaque ménage de l’étude, on retrouve une ou plusieurs lignes concernant la mère et sa consommation journalière et une ou plusieurs lignes concernant la consommation de l’enfant.*

mother\_chil\_baseline\_v1 <- bind\_rows(mother\_baseline\_v1,child\_baseline\_v1) %>% # empiler les deux bases   
 arrange(hhid,s1\_q2) # trié suivant le numéro du ménage et suivant la personne qui a consommé

- ii) Renommez toutes les variables de consommation energ\_kcal jusqu’à vit\_c\_mcg en ajoutant le suffixe \_b pour faire référence à l’enquête Baseline.

# Renommer les colonnes de consommation d'énergie et nutriments en ajoutant "\_b"  
  
mother\_chil\_baseline\_v1 <- mother\_chil\_baseline\_v1 %>%  
 rename\_with(~paste0(., "\_b"), # Renommer le nom de la variable par le nom concatener avec \_b  
 .cols = energ\_kcal:vit\_c\_mg) # spécifie les variables concernées par l'opération

- iii) Créez une base de données qui résume les consommations journalières totales par individu (somme des 4 repas) pour l’énergie et tous les nutriments en utilisant la commande merge.

# Calculer la somme pour chaque variable entre energ\_kcal\_b et vit\_c\_mg\_b  
base <- mother\_chil\_baseline\_v1 %>%  
 group\_by(hhid, s1\_q2) %>% ## Grouper suivant le numér du ménage et suivant le fait que ça soit mere ou enfant  
 summarise(across(energ\_kcal\_b:vit\_c\_mg\_b, \(x) sum(x, na.rm = TRUE)), .groups = "drop") # calculons la somme des consommation par individu et en ignorant les valeurs manquantes  
  
base # Afficher la base

## # A tibble: 2,128 × 11  
## hhid s1\_q2 energ\_kcal\_b protein\_g\_b lipid\_tot\_g\_b calcium\_mg\_b iron\_mg\_b  
## <chr> <dbl+l> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 494848… 1 [mot… 1570. 53.2 18.2 142. 21.4   
## 2 494848… 2 [chi… 1407. 50.1 17.4 172. 20.5   
## 3 494848… 1 [mot… 563. 11.7 6.49 116. 2.72  
## 4 494848… 2 [chi… 1001. 18.9 20.1 394. 6.51  
## 5 494848… 1 [mot… 1815. 79.9 7.71 189. 21.6   
## 6 494848… 2 [chi… 634. 27.4 3.59 51.3 8.81  
## 7 494848… 1 [mot… 6120. 205. 131. 936. 90.7   
## 8 494848… 2 [chi… 6965. 226. 132. 1204. 124.   
## 9 494848… 1 [mot… 2178. 71.8 11.0 201. 27.4   
## 10 494848… 2 [chi… 1062. 34.1 5.32 110. 12.9   
## # ℹ 2,118 more rows  
## # ℹ 4 more variables: zinc\_mg\_b <dbl>, vit\_b6\_mg\_b <dbl>, vit\_b12\_mcg\_b <dbl>,  
## # vit\_c\_mg\_b <dbl>

# Renommer les colonnes en supprimant le suffixe "\_e"  
names(base) <- gsub("\_e$", "", names(base))  
  
# Afficher les nouveaux noms des colonnes  
names(base)

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

- iv) Sauvegardez la base de données finale contenant seulement les hhid s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg iron\_mg zinc\_mg sous le nom « baseline\_final.dta ». Veuillez noter que pour cette base de données, pour chaque ménage, nous avons une ligne pour la mère et une ligne pour l’enfant.

base <- base %>%   
 select(hhid,s1\_q2,energ\_kcal\_b,protein\_g\_b,lipid\_tot\_g\_b, calcium\_mg\_b, iron\_mg\_b, zinc\_mg\_b )  
fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "baseline\_final.dta") # Création du dossier complet  
  
haven::write\_dta(base, fichier\_sauvegarde) # Sauvegarde

* \*i) Empilez les bases de données des mères (mother\_endline\_v1.dta) et des enfants (child\_endline\_v1.dta) de sorte que pour chaque ménage de l’étude, on retrouve une ou plusieurs lignes concernant la mère et sa consommation journalière et une ou plusieurs lignes concernant la consommation de l’enfant. Renommez toutes les variables de consommation energ\_kcal jusqu’à vit\_c\_mcg en ajoutant le suffixe \_e pour faire référence à l’enquête Endline.\*

mother\_chil\_endline\_v1 <- bind\_rows(mother\_endline\_v1.dta,child\_endline\_v1) %>% # empiler les deux bases   
 arrange(hhid,s1\_q2) # trié suivant le numéro du ménage et suivant la personne qui a consommé  
  
# Renommer les colonnes de consommation d'énergie et nutriments en ajoutant "\_b"  
  
mother\_chil\_endline\_v1 <- mother\_chil\_endline\_v1 %>%  
 rename\_with(~paste0(., "\_e"), # Renommer le nom de la variable par le nom concatener avec \_b  
 .cols = energ\_kcal:vit\_c\_mg) # spécifie les variables concernées par l'opération

- ii) Créez une base de données qui résume les consommations journalières totales par individu (somme des 4 repas) pour l’énergie et tous les nutriments en utilisant la commande collapse.

# Calculer la somme pour chaque variable entre energ\_kcal\_b et vit\_c\_mg\_b  
base <- mother\_chil\_endline\_v1 %>%  
 group\_by(hhid, s1\_q2) %>% ## Grouper suivant le numér du ménage et suivant le fait que ça soit mere ou enfant  
 summarise(across(energ\_kcal\_e:vit\_c\_mg\_e, \(x) sum(x, na.rm = TRUE)), .groups = "drop") # calculons la somme des consommation par individu et en ignorant les valeurs manquantes  
  
base # Afficher la base

## # A tibble: 2,128 × 11  
## hhid s1\_q2 energ\_kcal\_e protein\_g\_e lipid\_tot\_g\_e calcium\_mg\_e iron\_mg\_e  
## <chr> <dbl+l> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 494848… 1 [mot… 6992. 195. 284. 2804. 59.4   
## 2 494848… 2 [chi… 3960. 116. 149. 1718. 34.0   
## 3 494848… 1 [mot… 1292. 35.0 4.09 29.1 2.17   
## 4 494848… 2 [chi… 364. 9.81 1.13 7.95 0.572  
## 5 494848… 1 [mot… 3112. 73.9 124. 419. 37.8   
## 6 494848… 2 [chi… 3149. 107. 97.8 451. 51.8   
## 7 494848… 1 [mot… 4135. 66.6 123. 121. 27.2   
## 8 494848… 2 [chi… 3873. 66.7 118. 118. 27.9   
## 9 494848… 1 [mot… 2884. 109. 18.3 3197. 63.5   
## 10 494848… 2 [chi… 1208. 42.5 7.76 1159. 26.4   
## # ℹ 2,118 more rows  
## # ℹ 4 more variables: zinc\_mg\_e <dbl>, vit\_b6\_mg\_e <dbl>, vit\_b12\_mcg\_e <dbl>,  
## # vit\_c\_mg\_e <dbl>

# Renommer les colonnes en supprimant le suffixe "\_e"  
names(base) <- gsub("\_e$", "", names(base))  
  
# Afficher les nouveaux noms des colonnes  
names(base)

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

- iii) Sauvegarder la base de données finale contenant seulement les hhid s1\_q2 energ\_kcal protein\_g lipid\_tot\_g calcium\_mg iron\_mg zinc\_mg sous le nom « endline\_final.dta ». Veuillez noter que pour cette base de données, pour chaque ménage, nous avons une ligne pour la mère et une ligne pour l’enfant.

base <- base %>%   
 select(hhid,s1\_q2,energ\_kcal,protein\_g,lipid\_tot\_g, calcium\_mg, iron\_mg, zinc\_mg )  
fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "endline\_final.dta") # Création du dossier complet  
  
haven::write\_dta(base, fichier\_sauvegarde) # Sauvegarde

1. Fusionnez les données baseline\_final.dta et endline\_final.dta. La base de données résultant de cette fusion doit comporter le même nombre d’observations que les bases baseline\_final.dta et endline\_final.dta. Ainsi, sur une ligne d’observation, on retrouve à la fois les informations de l’enquête Baseline et de l’enquête Endline, concernant l’individu (mère ou enfant).

**Importation**

baseline\_final <- haven::read\_dta("../Données/baseline\_final.dta") # Importation de la base baseline\_final.dta  
  
endline\_final <- haven::read\_dta("../Données/endline\_final.dta") # Importation de la base endline\_final.dta

**Fusion**

base\_end <- left\_join(baseline\_final,endline\_final,by = c("hhid","s1\_q2"))

1. En utilisant la base base\_menage\_final, fusionnez les informations sur les caractéristiques des ménages telles que région, commune, village, statut matrimonial, taille du ménage, niveau d’éducation du chef de ménage, le ratio de dépendance, score HFIAS, la variable de traitement T1.

base <- base\_menage\_final %>%   
 select(hhid,regionid,communeid,villageid,hhsize,hh\_primary,dependencyratio,hfias\_score,T1)  
  
base\_end\_men <- full\_join(base\_end,base,by = "hhid")

1. Sauvegarder la base de données finale sous le nom base\_finale.dta.

fichier\_sauvegarde <- file.path(dossier\_sauvegarde, "base\_finale.dta.") # Création du dossier complet  
  
haven::write\_dta(base\_end\_men, fichier\_sauvegarde) # Sauvegarde

# Partie 3 : Statistiques descriptives avec le package gtsummary

**importation**

base\_final <- haven::read\_dta("../Données/base\_finale.dta") # Importation de la base base\_final.dta

# Résumé des variables quantitatives  
summary\_table\_quantitative <- base\_final %>%  
 select(energ\_kcal, protein\_g, lipid\_tot\_g, calcium\_mg) %>%  
 tbl\_summary(  
 statistic = list(all\_continuous() ~ "{mean} ({sd})") # Moyenne et écart-type  
 )%>%  
 modify\_header(label = "\*\*Taille de l'échantillon\*\*") %>%  
 bold\_labels() %>%   
   
as\_gt() %>%  
gt::tab\_header(  
 title = "Statistique sur la consommation des différents nutriments "  
 ) %>%  
gt::tab\_options(  
 heading.title.font.size = 18,  
 heading.title.font.weight = "black",  
 heading.background.color = "#0073C2"  
   
)  
  
summary\_table\_quantitative

Table 1: Statistique sur la consommation des différents nutriments

| **Taille de l'échantillon** | **N = 2,128***1* |
| --- | --- |
| **energ\_kcal** | 1,969 (1,440) |
| **protein\_g** | 60 (47) |
| **lipid\_tot\_g** | 39 (43) |
| **calcium\_mg** | 370 (478) |
| *1*Mean (SD) | |

base\_final%>%   
 select(s1\_q2,energ\_kcal\_b ,protein\_g\_b,lipid\_tot\_g\_b,calcium\_mg\_b)%>%   
 labelled::to\_factor() %>%  
 tbl\_summary( by = s1\_q2,  
 label=list(energ\_kcal\_b~"Consommation d'énergie ",  
 protein\_g\_b~"Consomation de protéine",  
 lipid\_tot\_g\_b~"Consomation de lipide ",  
 calcium\_mg\_b~"Consomation de calcium "),  
 digits=everything()~c(0,0,0,0),  
 statistic = list(everything()~"{mean} ({sd}) {median}"),  
 missing = "always",  
 missing\_text="Valeurs manquante"  
)%>%  
 modify\_header(label = "\*\*Taille de l'échantillon\*\*") %>%  
 bold\_labels() %>%   
   
as\_gt() %>%  
gt::tab\_header(  
 title = "Consomation "  
 ) %>%  
gt::tab\_options(  
 heading.title.font.size = 18,  
 heading.title.font.weight = "black",  
 heading.background.color = "#0073C2"  
   
)

Table 1: Consomation

| **Taille de l'échantillon** | **mother** N = 1,064*1* | **child** N = 1,064*1* | **mother and child** N = 0*1* |
| --- | --- | --- | --- |
| **Consommation d'énergie** | 2,407 (1,281) 2,116 | 1,500 (841) 1,310 | NA (NA) NA |
| Valeurs manquante | 0 | 0 | 0 |
| **Consomation de protéine** | 74 (74) 61 | 47 (38) 38 | NA (NA) NA |
| Valeurs manquante | 0 | 0 | 0 |
| **Consomation de lipide** | 39 (47) 25 | 29 (32) 19 | NA (NA) NA |
| Valeurs manquante | 0 | 0 | 0 |
| **Consomation de calcium** | 386 (381) 258 | 290 (301) 186 | NA (NA) NA |
| Valeurs manquante | 0 | 0 | 0 |
| *1*Mean (SD) Median | | | |