Test\_R

Dior MBENGUE

2025-02-25

# Installation des packages

#Vérification et installation des packages nécessaires  
packages <- c("haven", "utils", "dplyr", "tidyverse", "gtsummary", "survey")  
for (pkg in packages) {  
 if (!require(pkg, character.only = TRUE)) install.packages(pkg, dependencies = TRUE)  
 library(pkg, character.only = TRUE)  
}

## Le chargement a nécessité le package : haven

## Le chargement a nécessité le package : dplyr

##   
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':  
##   
## filter, lag

## Les objets suivants sont masqués depuis 'package:base':  
##   
## intersect, setdiff, setequal, union

## Le chargement a nécessité le package : tidyverse

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ readr 2.1.5  
## ✔ ggplot2 3.5.1 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.4 ✔ tibble 3.2.1  
## ✔ purrr 1.0.4 ✔ tidyr 1.3.1  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors  
## Le chargement a nécessité le package : gtsummary  
##   
## Le chargement a nécessité le package : survey  
##   
## Le chargement a nécessité le package : grid  
##   
## Le chargement a nécessité le package : Matrix  
##   
##   
## Attachement du package : 'Matrix'  
##   
##   
## Les objets suivants sont masqués depuis 'package:tidyr':  
##   
## expand, pack, unpack  
##   
##   
## Le chargement a nécessité le package : survival  
##   
##   
## Attachement du package : 'survey'  
##   
##   
## L'objet suivant est masqué depuis 'package:graphics':  
##   
## dotchart

#importation des bases

men<-haven::read\_dta("../data/base\_menage.dta")  
fc\_baseline<-haven::read\_dta("../data/food\_comp\_child\_baseline.dta")  
fc\_endline<-haven::read\_dta("../data/food\_comp\_child\_endline.dta")  
fm\_baseline<-haven::read\_dta("../data/food\_comp\_mother\_baseline.dta")  
fm\_endline<-haven::read\_dta("../data/food\_comp\_mother\_endline.dta")

#Vérification de l’existence de doublon

#base ménage  
sum(duplicated(men))

## [1] 1

head(duplicated(men), 1)

## [1] FALSE

#base food\_comp\_child\_baseline  
sum(duplicated(fc\_baseline))

## [1] 0

head(duplicated(fc\_baseline), 1)

## [1] FALSE

#base food\_comp\_child\_endline  
sum(duplicated(fc\_endline))

## [1] 0

head(duplicated(fc\_endline), 1)

## [1] FALSE

#base food\_comp\_mother\_baseline  
  
sum(duplicated(fm\_baseline))

## [1] 0

head(duplicated(fm\_baseline), 1)

## [1] FALSE

# base food\_comp\_mother\_endline  
sum(duplicated(fm\_endline))

## [1] 0

head(duplicated(fm\_endline), 1)

## [1] FALSE

#Supression des doublons dans la base ménage

men <- men[!duplicated(men), ]

# vérification de la cohérence des noms des variables

colnames(fc\_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"

colnames(fc\_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"

colnames(fm\_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"

colnames(fm\_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"

#Rectification des noms des variables

colnames(fm\_baseline)[colnames(fm\_baseline) == "V1"] <- "energ\_kcal"  
colnames(fm\_baseline)[colnames(fm\_baseline) == "V9"] <- "zinc\_mg"  
colnames(fm\_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"

#Vérification de présences de données manquantes

# Calculer le nombre de valeurs manquantes dans toute la base de données  
  
sum(is.na(men))

## [1] 1

sum(is.na(fc\_baseline))

## [1] 3474

sum(is.na(fc\_endline))

## [1] 4788

sum(is.na(fm\_baseline))

## [1] 6069

sum(is.na(fm\_endline))

## [1] 5976

Consommation d’énergie moyenne pour l’ensembles des mères

mean(fm\_baseline$energ\_kcal, na.rm = TRUE)

## [1] 714.7165

les mères consomment en moyenne 714.7165 calories pendant chaque repas

#Sauvegarde des bases finales

haven::write\_dta(fm\_baseline, "mother\_baseline\_v1.dta")  
haven::write\_dta(fm\_endline, "mother\_endline\_v1.dta")  
haven::write\_dta(fc\_baseline, "child\_baseline\_v1.dta")  
haven::write\_dta(fc\_endline, "child\_endline\_v1.dta")  
haven::write\_dta(men, "base\_menage\_final.dta")

# chargement des nouvelles bases baseline

# Charger les bases de données  
menf <- haven::read\_dta("base\_menage\_final.dta")  
cbv1<- haven::read\_dta("child\_baseline\_v1.dta")  
mbv1<- haven::read\_dta("mother\_baseline\_v1.dta")

# Empilement pour les baseline

men\_merge <- merge(menf, mbv1, by = "hhid", all = TRUE)  
men\_mother\_child\_merge <- merge(men\_merge, cbv1, by = "hhid", all = TRUE)  
  
  
## j'ai mergé les bases child\_baseline\_v1.dta et mother\_baseline\_v1.dta dans la base base\_menage\_final.dta

# Vérification de la base mergée

colnames(men\_mother\_child\_merge)

## [1] "hhid" "regionid.x" "communeid.x" "villageid.x"   
## [5] "hhsize" "poly" "hh\_primary" "s1\_q2.x"   
## [9] "s1\_q4a" "s2\_q1" "s2\_q2" "s2\_q4"   
## [13] "s29\_q1" "demgrp1" "demgrp2" "demgrp3"   
## [17] "demgrp4" "demgrp5" "dependencyratio" "hfias\_score"   
## [21] "T1" "regionid.y" "communeid.y" "villageid.y"   
## [25] "round.x" "s1\_q0.x" "s1\_q1.x" "s1\_q2.y"   
## [29] "energ\_kcal.x" "protein\_g.x" "lipid\_tot\_g.x" "calcium\_mg.x"   
## [33] "iron\_mg.x" "zinc\_mg.x" "vit\_b6\_mg.x" "vit\_b12\_mcg.x"   
## [37] "vit\_c\_mg.x" "regionid" "communeid" "villageid"   
## [41] "round.y" "s1\_q0.y" "s1\_q1.y" "s1\_q2"   
## [45] "energ\_kcal.y" "protein\_g.y" "lipid\_tot\_g.y" "calcium\_mg.y"   
## [49] "iron\_mg.y" "zinc\_mg.y" "vit\_b6\_mg.y" "vit\_b12\_mcg.y"   
## [53] "vit\_c\_mg.y"

# Renommer les noms des variables de energ\_kcal à vit\_c\_mcg en ajoutant \_b

# Liste des colonnes à renommer  
cols\_to\_rename <- c("energ\_kcal.x", "protein\_g.x", "lipid\_tot\_g.x", "calcium\_mg.x", "iron\_mg.x","zinc\_mg.x","vit\_b6\_mg.x","vit\_b12\_mcg.x","vit\_c\_mg.x")  
  
# Appliquer le suffixe "\_b"  
men\_mother\_child\_merge <- men\_mother\_child\_merge %>%  
 rename\_with(~ paste0(.x, "\_b"), .cols = all\_of(cols\_to\_rename))

#chargement des nouvelles bases endline

cev1<- haven::read\_dta("child\_endline\_v1.dta")  
mev1<- haven::read\_dta("mother\_endline\_v1.dta")

# chargement des bases endline

men\_merge\_end <- merge(menf, mev1, by = "hhid", all = TRUE)  
men\_mother\_child\_merge\_end <- merge(men\_merge, cev1, by = "hhid", all = TRUE)

# vérification de la base mergée endline

colnames(men\_mother\_child\_merge\_end)

## [1] "hhid" "regionid.x" "communeid.x" "villageid.x"   
## [5] "hhsize" "poly" "hh\_primary" "s1\_q2.x"   
## [9] "s1\_q4a" "s2\_q1" "s2\_q2" "s2\_q4"   
## [13] "s29\_q1" "demgrp1" "demgrp2" "demgrp3"   
## [17] "demgrp4" "demgrp5" "dependencyratio" "hfias\_score"   
## [21] "T1" "regionid.y" "communeid.y" "villageid.y"   
## [25] "round.x" "s1\_q0.x" "s1\_q1.x" "s1\_q2.y"   
## [29] "energ\_kcal.x" "protein\_g.x" "lipid\_tot\_g.x" "calcium\_mg.x"   
## [33] "iron\_mg.x" "zinc\_mg.x" "vit\_b6\_mg.x" "vit\_b12\_mcg.x"   
## [37] "vit\_c\_mg.x" "regionid" "communeid" "villageid"   
## [41] "round.y" "s1\_q0.y" "s1\_q1.y" "s1\_q2"   
## [45] "energ\_kcal.y" "protein\_g.y" "lipid\_tot\_g.y" "calcium\_mg.y"   
## [49] "iron\_mg.y" "zinc\_mg.y" "vit\_b6\_mg.y" "vit\_b12\_mcg.y"   
## [53] "vit\_c\_mg.y"

# on constate que les colonnes des base mères et enfants ont été ajouté dans la base ménage

#Statistiques descriptives ## Statistique univariée

men\_mother\_child\_merge\_end %>% labelled::to\_factor() %>%  
 select(regionid, villageid, hhsize, energ\_kcal.y, energ\_kcal.x, s2\_q1, s2\_q2, s2\_q4, poly, protein\_g.x, protein\_g.y) %>%  
 tbl\_summary(  
 label = list(  
 regionid ~ "Répartition par région de la mère",  
 villageid ~ "Répartition par village de la mère",  
 hhsize ~ "Taille du ménage",  
 energ\_kcal.x ~ "Energie consommée de la mère",  
 energ\_kcal.y ~ "Energie consommée de l'enfant",  
 s2\_q1 ~ "chef de ménage alphabétise en la langue locale",  
 s2\_q2 ~ "chef de ménage alphabétise en français",  
 s2\_q4 ~ "chef de ménage a été scolarisé",  
 poly ~ "chef de ménage polygame",  
 protein\_g.x ~ "protéine\_g de la mère",  
 protein\_g.y ~ "protéine\_g de l'enfant"  
 ),  
 statistic = list(all\_categorical() ~ "{n}({p}%)"),  
 missing = "always",  
 missing\_text = "Les valeurs manquantes"  
 ) %>% modify\_header(label = "Statistiques sur les variables de la base mergée")

| Statistiques sur les variables de la base mergée | **N = 17,024***1* |
| --- | --- |
| Répartition par région de la mère |  |
| 1 | 3,008(18%) |
| 2 | 9,904(58%) |
| 3 | 4,112(24%) |
| Les valeurs manquantes | 0 |
| Répartition par village de la mère | 1,255 (1,002, 1,548) |
| Les valeurs manquantes | 0 |
| Taille du ménage | 8.0 (6.0, 11.0) |
| Les valeurs manquantes | 0 |
| Energie consommée de l'enfant | 349 (229, 547) |
| Les valeurs manquantes | 2,128 |
| Energie consommée de la mère | 622 (413, 899) |
| Les valeurs manquantes | 2,692 |
| chef de ménage alphabétise en la langue locale |  |
| Non | 15,056(88%) |
| Oui | 1,968(12%) |
| Les valeurs manquantes | 0 |
| chef de ménage alphabétise en français |  |
| Non | 13,440(79%) |
| Oui | 3,584(21%) |
| Les valeurs manquantes | 0 |
| chef de ménage a été scolarisé |  |
| Non | 12,048(71%) |
| Oui | 4,976(29%) |
| Les valeurs manquantes | 0 |
| chef de ménage polygame |  |
| Non | 8,864(52%) |
| Oui | 8,160(48%) |
| Les valeurs manquantes | 0 |
| protéine\_g de la mère | 17 (11, 26) |
| Les valeurs manquantes | 2,692 |
| protéine\_g de l'enfant | 10 (6, 17) |
| Les valeurs manquantes | 2,128 |
| *1*n(%); Median (Q1, Q3) | |

# interprétation

##Composition du ménage et répartition géographique

La répartition géographique montre une forte concentration des individus dans la Région 2 (58%), suivie des Régions 1 (18%) et 3 (24%). Cela suggère que la majorité des mères et enfants viennent de cette région, ce qui peut avoir des implications sur l’accès aux ressources, à la nutrition et aux services de santé. La taille du ménage est relativement grande, avec une médiane de 8 membres, ce qui indique que de nombreux ménages sont composés de plusieurs générations, notamment avec des enfants, des parents et parfois des grands-parents. La distribution est relativement homogène, car l’IQR va de 6 à 11 membres.

## Consommation énergétique des mères et des enfants

La consommation énergétique de l’enfant (349 kcal) est globalement inférieure à celle de la mère (622 kcal). Cela peut refléter les différences de besoins énergétiques entre les groupes d’âge, mais aussi la disponibilité de nourriture et d’autres facteurs nutritionnels. Les valeurs manquantes dans les variables de consommation énergétique (en particulier pour la mère et l’enfant) sont relativement élevées (2,692 pour la mère et 2,128 pour l’enfant), ce qui suggère qu’il y a des lacunes dans la collecte ou la disponibilité des données sur l’alimentation.

## Alphabétisation et éducation du chef de ménage

Les chefs de ménage montrent des taux élevés d’analphabétisme, avec une majorité n’étant pas alphabétisée en langue locale (88%) ni en français (79%). Cependant, environ 12% des chefs de ménage sont alphabétisés en langue locale, et 21% en français. Cela peut avoir des implications sur l’accès à l’information, à l’éducation et à la participation aux décisions de santé ou nutritionnelles. De plus, 29% des chefs de ménage ont été scolarisés, ce qui est relativement faible, indiquant peut-être un défi dans l’accès à une éducation formelle pour une proportion significative de la population.

## Polygamie et structure familiale

La répartition sur la polygamie montre que presque 48% des chefs de ménage sont polygames. Cela peut refléter des normes culturelles ou sociales spécifiques aux régions étudiées. La polygamie peut influencer la structure familiale, la répartition des ressources et les dynamiques au sein du ménage.

## Protéines consommées

Les données sur la consommation de protéines (17 g pour la mère et 10 g pour l’enfant) montrent une tendance vers une consommation relativement modeste de protéines, avec des intervalles interquartiles indiquant une variabilité dans les habitudes alimentaires. Toutefois, les valeurs manquantes pour la consommation de protéines (en particulier pour la mère) sont également assez élevées, ce qui peut nuire à l’exactitude de l’analyse nutritionnelle globale.