

Statistiques descriptives

ONANENA AMANA Jeanne De La Flèche

2025-02-11

```
library(gtsummary)
library(labelled)
library(haven)
library(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
```

```
library(gt)
```

```
#chargeons les bases de données
data1 <- read_dta("ehcvm_menage_bfa2021.dta")
data2 <- read_dta("ehcvm_welfare_2b_bfa2021.dta")
```

```
names(data1)
```

```
## [1] "country" "hhid" "year" "grappe" "menage"
## [6] "vague" "logem" "mur" "toit" "sol"
## [11] "eauboi_ss" "eauboi_sp" "elec_ac" "elec_ur" "elec_ua"
## [16] "ordure" "toilet" "eva_toi" "eva_eau" "tv"
## [21] "fer" "frigo" "cuisin" "ordin" "decod"
## [26] "car" "superf" "grosrum" "petitrum" "porc"
## [31] "lapin" "volail" "sh_id_demo" "sh_co_natu" "sh_co_eco"
## [36] "sh_id_eco" "sh_co_vio" "sh_co_oth" "enquete"
```

```
data1 %>% select(logem, toit,sol) %>% tbl_summary()
```

```
## ! Column(s) "logem", "toit", and "sol" are class "haven_labelled".
## i This is an intermediate datastructure not meant for analysis.
## i Convert columns with `haven::as_factor()`, `labelled::to_factor()`,
##   `labelled::unlabelled()`, and `unclass()`. Failure to convert may have
##   unintended consequences or result in error.
## <https://haven.tidyverse.org/articles/semantics.html>
## <https://larmarange.github.io/labelled/articles/intro\_labelled.html#unlabelled>
```

Le tableau précédent ne donne aucune information. Nous devons l'améliorer pour une interprétabilité.

```
###les tableaux doivent être explicites. Il est donc important de toujours labelliser les variables
data1 %>% to_factor()%>%select(logem, toit,sol) %>%tbl_summary()
```

Characteristic	N = 7,176 ^I
Occupation logement	
1	1,669 (23%)
2	3,733 (52%)
3	962 (13%)
4	812 (11%)
toit en materiaux definitifs	
0	613 (8.5%)
1	6,563 (91%)
Sol en materiaux definitifs	
0	1,383 (19%)
1	5,793 (81%)
^I n (%)	

Characteristic	N = 7,176 ^I
Occupation logement	
Proprietaire titre	1,669 (23%)
Proprietaire sans titre	3,733 (52%)
Locataire	962 (13%)
Autre	812 (11%)
toit en materiaux definitifs	
Non	613 (8.5%)
Oui	6,563 (91%)
Sol en materiaux definitifs	
Non	1,383 (19%)
Oui	5,793 (81%)
^I n (%)	

```
###lDescription
data1 %>%
  to_factor() %>% # Conversion des variables catégorielles en facteurs
  select(logem, toit, sol) %>%
  tbl_summary(
    label = list(
      logem ~ "Type de logement du chef de ménage",
      toit ~ "Type de toit",
      sol ~ "Type de sol"
    )
  ) %>% modify_header(label ~ "Caractéristiques")
```

```
###lDescription
data1 %>%
  to_factor() %>% # Convertir les variables catégorielles en facteurs
  select(logem, toit, sol, superf, grosrum) %>%
```

Caractéristiques	N = 7,176 ^I
Type de logement du chef de ménage	
Propriétaire titre	1,669 (23%)
Propriétaire sans titre	3,733 (52%)
Locataire	962 (13%)
Autre	812 (11%)
Type de toit	
Non	613 (8.5%)
Oui	6,563 (91%)
Type de sol	
Non	1,383 (19%)
Oui	5,793 (81%)
^I n (%)	

Characteristic	N = 7,176 ^I
Type de logement du chef de ménage	
Propriétaire titre	1,669 (23%)
Propriétaire sans titre	3,733 (52%)
Locataire	962 (13%)
Autre	812 (11%)
Type de toit	
Non	613 (8.5%)
Oui	6,563 (91%)
Type de sol	
Non	1,383 (19%)
Oui	5,793 (81%)
superficie	3.46(3.84)
Unknown	2,817
Nombre de gros ruminants	1.74(6.81)
^I n (%); Mean(SD)	

```
tbl_summary(
  label = list(
    logem ~ "Type de logement du chef de ménage",
    toit ~ "Type de toit",
    sol ~ "Type de sol",
    superf ~ "superficie",
    grosrum ~ "Nombre de gros ruminants"
  ),
  statistic = list(grosrum ~ "{mean}({sd})", superf ~ "{mean}({sd})")
)
```

```
# Description
data1 %>%
```

Characteristic	N = 7,176 ¹
Superficie agricole en hectare	3.46(3.84)
Unknown	2,817
Nbr gros ruminants	1.74(6.81)
¹ Mean(SD)	

Characteristic	N = 7,176 ¹
Superficie agricole en hectare	3(4)
Unknown	2,817
Nbr gros ruminants	2(7)
¹ Mean(SD)	

Characteristic	N = 7,176 ¹
Superficie agricole en hectare	3(4)
Missing	2,817
Nbr gros ruminants	2(7)
Missing	0
¹ Mean(SD)	

```

to_factor() %>%
select(superf,grosum) %>%
tbl_summary(
  statistic = list(grosum~ "{mean}({sd})", superf~ "{mean}({sd})")
)

# Description
data1 %>%
  to_factor() %>%
  select(superf,grosum) %>%
  tbl_summary(
    statistic = list(grosum~ "{mean}({sd})", superf~ "{mean}({sd})"),digits = everything()~c(0,0) #even
  )

# Description
# missing = always permet de toujours afficher les valeurs manquantes
data1 %>%
  to_factor() %>%
  select(superf,grosum) %>%
  tbl_summary(
    statistic = list(grosum~ "{mean}({sd})", superf~ "{mean}({sd})"),digits = everything()~c(0,0),
    missing = "always",
    missing_text= "Missing",
  )

```

```

data2 %>%
  labelled::to_factor() %>%
  select(hgender, hage, hmstat, heduc, hdiploma, hcsp) %>%
  tbl_summary(
    label = list(
      hgender = "Genre du chef du ménage",
      hage = "Âge du chef du ménage",
      hmstat = "Statut matrimonial",
      heduc = "Niveau d'éducation",
      hdiploma = "Diplôme obtenu",
      hcsp = "Catégorie socio-professionnelle"
    ),
    missing = "ifany",
    statistic = list(
      all_categorical() ~ "{n} ({p}%)",
      all_continuous() ~ "{mean} ({sd})"
    )
  ) %>% modify_header(label ~ "Caractéristiques")

```

```

tableau <- data2 %>%
  labelled::to_factor() %>%
  select(hgender, hage, hmstat, heduc, hdiploma, hcsp) %>%
  tbl_summary(
    label = list(
      hgender = "Genre du chef du ménage",
      hage = "Âge du chef du ménage",
      hmstat = "Statut matrimonial",
      heduc = "Niveau d'éducation",
      hdiploma = "Diplôme obtenu",
      hcsp = "Catégorie socio-professionnelle"
    ),
    missing = "ifany", # Affichons les valeurs manquantes si elles existent
    statistic = list(
      all_categorical() ~ "{n} ({p}%)", # Pour les variables qualitatives
      all_continuous() ~ "{mean} ({sd})" # Pour les variables quantitatives
    )
  ) %>% modify_header(label ~ "Caractéristiques") %>%
  as_gt() %>% # Conversion en tableau avec le package gt
  tab_style(
    style = list(cell_fill(color = "lightblue")),
    locations = cells_column_labels()
  )

# Affichons tableau dans le Viewer
tableau

```

Caractéristiques	N = 7,176 ^I
Genre du chef du ménage	
Masculin	6,101 (85%)
Féminin	1,075 (15%)
Âge du chef du ménage	46 (15)
Statut matrimonial	
Célibataire	349 (4.9%)
Marié(e) monogame	4,352 (61%)
Marié(e) polygame	1,572 (22%)
Union libre	199 (2.8%)
Veuf(ve)	568 (7.9%)
Divorcé(e)	50 (0.7%)
Séparé(e)	86 (1.2%)
Niveau d'éducation	
Aucun	4,706 (66%)
Maternelle	0 (0%)
Primaire	1,051 (15%)
Postprimaire général	622 (8.7%)
Postprimaire technique	8 (0.1%)
Secondaire général	395 (5.5%)
Secondaire technique	39 (0.5%)
Postsecondaire	44 (0.6%)
Superieur	311 (4.3%)
Diplôme obtenu	
Aucun	5,551 (77%)
CEPE	726 (10%)
BEPC	394 (5.5%)
CAP	20 (0.3%)
BT	5 (<0.1%)
BAC	229 (3.2%)
DEUG, DUT, BTS	58 (0.8%)
Licence	110 (1.5%)
Maitrise	43 (0.6%)
Master/DEA/DESS	36 (0.5%)
Doctorat/Phd	4 (<0.1%)
Catégorie socio-professionnelle	
Cadre supérieur	76 (1.2%)
Cadre moyen/agent de maîtrise	398 (6.4%)
Ouvrier ou employé qualifié	358 (5.7%)
Ouvrier ou employé non qualifié	494 (7.9%)
Manœuvre, aide ménagère	93 (1.5%)
Stagiaire ou Apprenti rémunéré	82 (1.3%)
Stagiaire ou Apprenti non rémunéré	21 (0.3%)
Travailleur Familial contribuant pour une entreprise familial	3 (<0.1%)
Travailleur pour compte propre	4,613 (74%)
Patron	100 (1.6%)
Unknown	938

^I n (%); Mean (SD)

Caractéristiques	N = 7,176 ^I
Genre du chef du ménage	
Masculin	6,101 (85%)
Féminin	1,075 (15%)
Âge du chef du ménage	46 (15)
Statut matrimonial	
Célibataire	349 (4.9%)
Marié(e) monogame	4,352 (61%)
Marié(e) polygame	1,572 (22%)
Union libre	199 (2.8%)
Veuf(ve)	568 (7.9%)
Divorcé(e)	50 (0.7%)
Séparé(e)	86 (1.2%)
Niveau d'éducation	
Aucun	4,706 (66%)
Maternelle	0 (0%)
Primaire	1,051 (15%)
Postprimaire général	622 (8.7%)
Postprimaire technique	8 (0.1%)
Secondaire général	395 (5.5%)
Secondaire technique	39 (0.5%)
Postsecondaire	44 (0.6%)
Superieur	311 (4.3%)
Diplôme obtenu	
Aucun	5,551 (77%)
CEPE	726 (10%)
BEPC	394 (5.5%)
CAP	20 (0.3%)
BT	5 (<0.1%)
BAC	229 (3.2%)
DEUG, DUT, BTS	58 (0.8%)
Licence	110 (1.5%)
Maitrise	43 (0.6%)
Master/DEA/DESS	36 (0.5%)
Doctorat/Phd	4 (<0.1%)
Catégorie socio-professionnelle	
Cadre supérieur	76 (1.2%)
Cadre moyen/agent de maîtrise	398 (6.4%)
Ouvrier ou employé qualifié	358 (5.7%)
Ouvrier ou employé non qualifié	494 (7.9%)
Manœuvre, aide ménagère	93 (1.5%)
Stagiaire ou Apprenti rémunéré	82 (1.3%)
Stagiaire ou Apprenti non rémunéré	21 (0.3%)
Travailleur Familial contribuant pour une entreprise familial	3 (<0.1%)
Travailleur pour compte propre	4,613 (74%)
Patron	100 (1.6%)
Unknown	938

^I n (%); Mean (SD)