

---

title: “TESTS STATISTIQUES SUR LES CONSOMMATIONS ALIMENTAIRES EN ZONE RURALE ET URBAINE”

author: “RIRADJIM NGARMOUNDOU Trésor”

date: “15 May 2025”

output:

pdf\_document:

toc: true

number\_sections: true

---

```
# Définir chemins
zip_path <- "/Users/HP/Downloads/SEN2018_menage.zip"
out_dir <- "/Users/HP/Downloads/SEN2018_menage"
if (!dir.exists(out_dir) && file.exists(zip_path)) unzip(zip_path, exdir = out_dir)
dta_files <- list.files(out_dir, pattern = "\\\\.dta$", full.names = TRUE, recursive = TRUE)
data_list <- setNames(lapply(dta_files, read_dta), basename(dta_files))

# Création des données individuelles
roster <- data_list[["s01_me_SEN2018.dta"]] %>%
  mutate(id = paste(grappe, menage, sep = "_"))

# Données de consommation totales déjà préparées
cons_food <- data_list[["s08b1_me_SEN2018.dta"]] %>%
  mutate(id = paste(grappe, menage, sep = "_"))

food_vars <- grep("^s08b02", names(cons_food), value = TRUE)
food_totals <- cons_food %>%
  select(id, all_of(food_vars)) %>%
  mutate(food_total = rowSums(across(all_of(food_vars)), na.rm = TRUE)) %>%
  select(id, food_total) %>%
  filter(food_total > 0) %>%
  distinct(id, .keep_all = TRUE)

zone_df <- data_list[["s00_me_SEN2018.dta"]] %>%
  mutate(id = paste(grappe, menage, sep = "_")) %>%
  select(id, s00q04) %>%
  rename(zone_code = s00q04) %>%
  mutate(zone = ifelse(zone_code == 1, "Urbaine", "Rurale")) %>%
  select(id, zone)

# Effectif par ménage + zone + conso
indiv_df <- roster %>%
  group_by(id) %>% summarise(n_indiv = n(), .groups = "drop")

df_indiv <- indiv_df %>%
  inner_join(food_totals, by = "id") %>%
  inner_join(zone_df, by = "id") %>%
  mutate(food_per_indiv = food_total / n_indiv)
```

## 1. Contexte

Ce rapport examine la **consommation alimentaire par individu** selon :

- la **taille du ménage**,
- la **zone de résidence** (urbaine/rurale),
- un **modèle multivarié** intégrant les deux facteurs.

## 2. Analyse descriptive

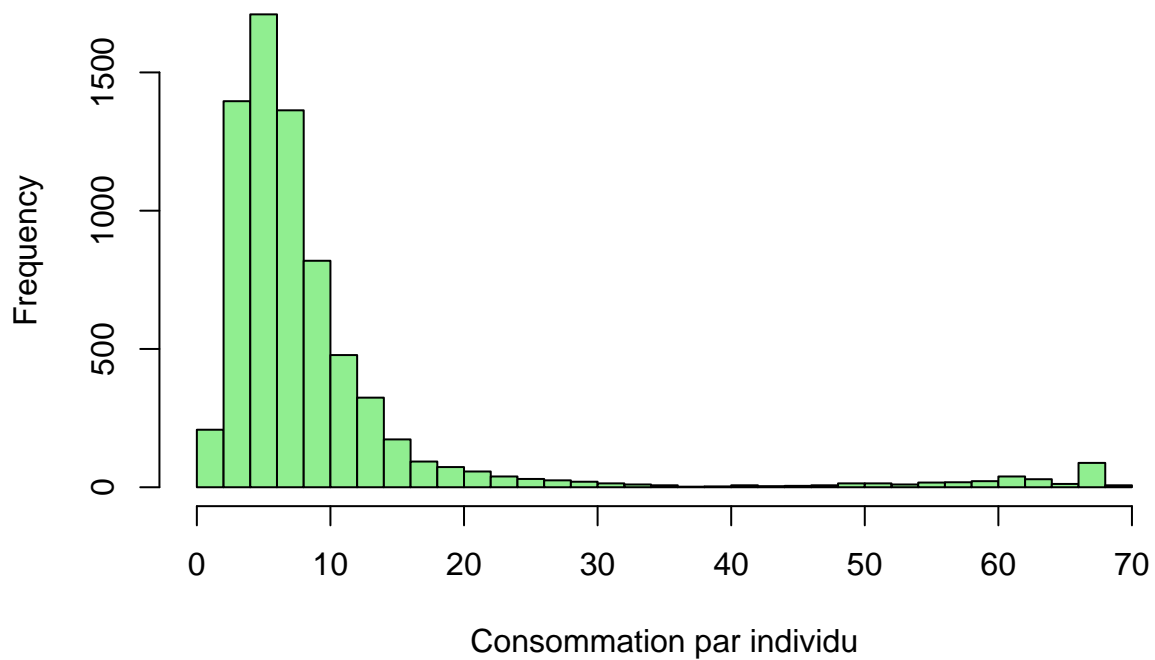
```
desc_zone <- df_indiv %>%  
  group_by(zone) %>%  
  summarise(  
    n_menages = n(),  
    mean_food = mean(food_per_indiv),  
    sd_food = sd(food_per_indiv)  
  )  
knitr::kable(desc_zone, caption = "Statistiques descriptives de la consommation individuelle par zone")
```

Table 1: Statistiques descriptives de la consommation individuelle par zone

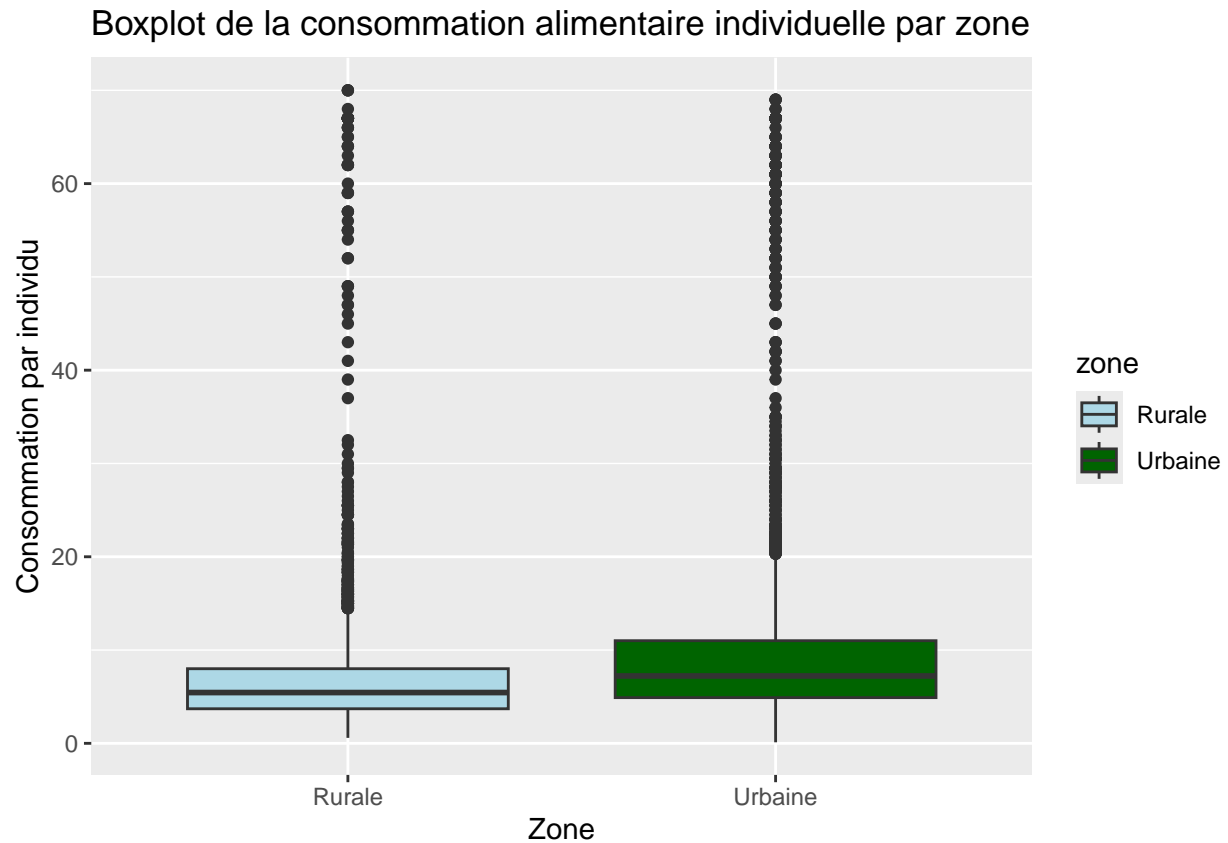
zone	n_menages	mean_food	sd_food
Rurale	3212	7.391587	8.285034
Urbaine	3925	11.384799	13.532003

```
hist(df_indiv$food_per_indiv, breaks = 30, col = "lightgreen",  
     main = "Distribution de la consommation alimentaire individuelle",  
     xlab = "Consommation par individu")
```

## Distribution de la consommation alimentaire individuelle

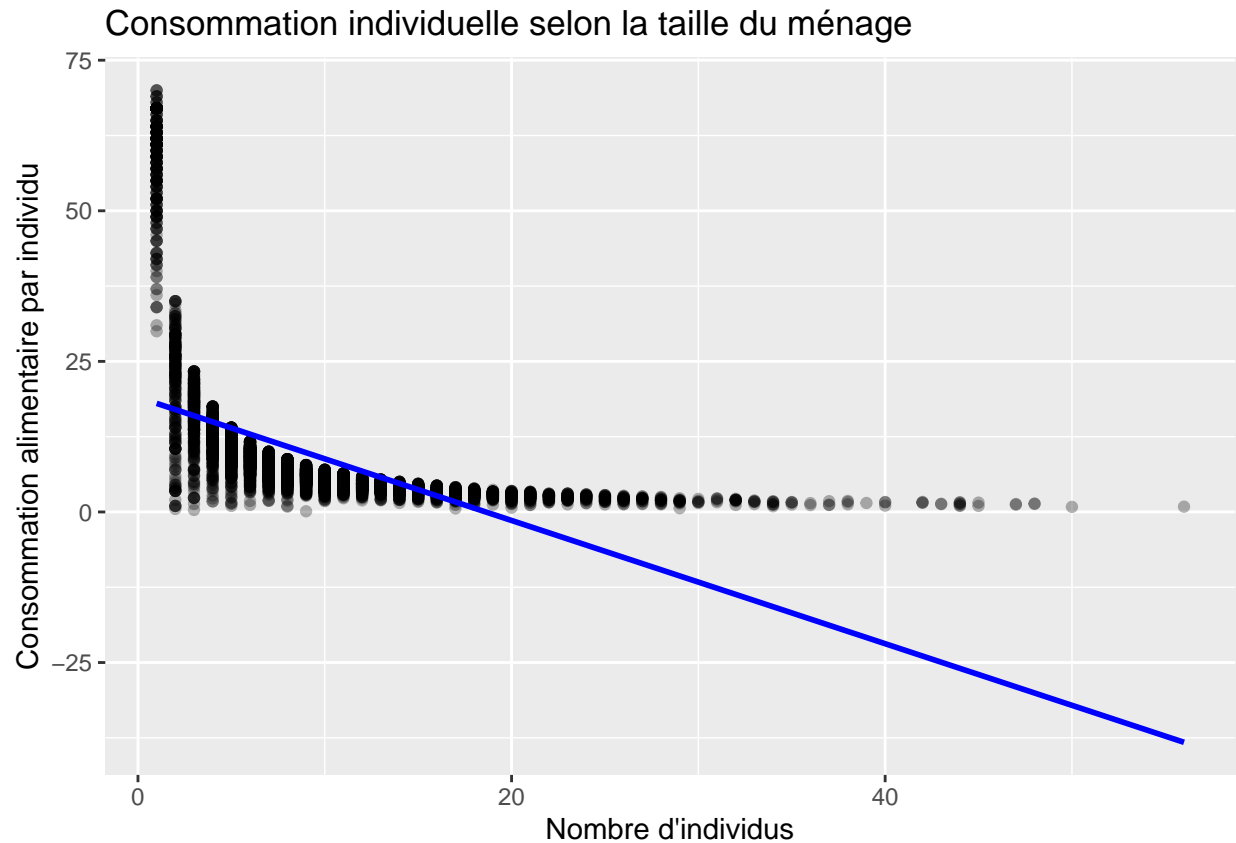


```
ggplot(df_indiv, aes(x = zone, y = food_per_indiv, fill = zone)) +  
  geom_boxplot() +  
  labs(title = "Boxplot de la consommation alimentaire individuelle par zone",  
        x = "Zone", y = "Consommation par individu") +  
  scale_fill_manual(values = c("Urbaine" = "darkgreen", "Rurale" = "lightblue"))
```



### 3. Relation entre taille du ménage et consommation

```
ggplot(df_indiv, aes(x = n_indiv, y = food_per_indiv)) +  
  geom_point(alpha = 0.3) +  
  geom_smooth(method = "lm", se = FALSE, color = "blue") +  
  labs(title = "Consommation individuelle selon la taille du ménage",  
        x = "Nombre d'individus", y = "Consommation alimentaire par individu")
```



#### 4. Test de corrélation

```
cor.test(df_indiv$n_indiv, df_indiv$food_per_indiv, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: df_indiv$n_indiv and df_indiv$food_per_indiv
## S = 1.1538e+11, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.9042487
```

**Interprétation :** la corrélation de Spearman évalue le lien monotone entre la taille du ménage et la consommation individuelle. La présence de valeurs identiques (ties) rend la p-value approximative.

## 5. Comparaison entre zones (urbaine vs rurale)

### Test de normalité (Shapiro-Wilk)

```
shapiro.test(df_indiv$food_per_indiv[df_indiv$zone == "Urbaine"])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: df_indiv$food_per_indiv[df_indiv$zone == "Urbaine"]  
## W = 0.56338, p-value < 2.2e-16
```

```
shapiro.test(df_indiv$food_per_indiv[df_indiv$zone == "Rurale"])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: df_indiv$food_per_indiv[df_indiv$zone == "Rurale"]  
## W = 0.49368, p-value < 2.2e-16
```

Les deux p-values étant  $< 0.05$ , on rejette l'hypothèse de normalité  $\rightarrow$  pas de test t.

### Homogénéité des variances (Levene)

```
leveneTest(food_per_indiv ~ zone, data = df_indiv)
```

```
## Levene's Test for Homogeneity of Variance (center = median)  
##           Df F value    Pr(>F)  
## group      1 127.42 < 2.2e-16 ***  
##           7135  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hétérogénéité détectée : les variances sont significativement différentes selon les zones.

### Test de Wilcoxon (non-paramétrique)

```
wilcox.test(food_per_indiv ~ zone, data = df_indiv)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: food_per_indiv by zone  
## W = 4628035, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0
```

Ce test indique si la consommation individuelle médiane diffère entre zones.

## Taille d'effet (Cohen's d)

```
cohens_d(df_indiv, food_per_indiv ~ zone)
```

```
## # A tibble: 1 x 7
##   .y.      group1 group2 effsize    n1    n2 magnitude
## * <chr>      <chr>  <chr>    <dbl> <int> <int> <ord>
## 1 food_per_indiv Rurale Urbaine  -0.356  3212  3925 small
```

Interprétation de la magnitude :

- $< 0.2$  : négligeable
- $0.2 - 0.5$  : petit effet
- $0.5 - 0.8$  : moyen
- $0.8$  : grand effet

## 6. Modèle multivarié

```
model <- lm(log(food_per_indiv + 1) ~ zone + n_indiv, data = df_indiv)
summary(model)
```

```
##
## Call:
## lm(formula = log(food_per_indiv + 1) ~ zone + n_indiv, data = df_indiv)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.37944 -0.19898 -0.06222  0.10666  2.50878
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.7684479  0.0109169  253.59  <2e-16 ***
## zoneUrbaine  0.1477093  0.0097025   15.22  <2e-16 ***
## n_indiv      -0.0830111  0.0008116 -102.29  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4035 on 7134 degrees of freedom
## Multiple R-squared:  0.6147, Adjusted R-squared:  0.6146
## F-statistic: 5690 on 2 and 7134 DF, p-value: < 2.2e-16
```

Le modèle linéaire estime l'impact combiné de la zone et de la taille du ménage sur la consommation alimentaire individuelle.

## 7. Conclusion

- Les tailles de ménage influencent négativement la consommation individuelle.
- La zone urbaine est associée à une consommation significativement plus élevée.
- Le modèle explique plus de 60% de la variance ( $R^2 > 0.6$ ).
- Les données ne sont pas normales  $\rightarrow$  test non paramétrique utilisé.