

RAPPORT D'ANALYSE DE TESTS STATISTIQUES SUR LES CONSOMMATIONS ALIMENTAIRES EN ZONE URBAINE ET RURALE

RIRADJIM NGARMOUNDOU Trésor

2025-05-15

Contents

0.1	Contexte	2
0.2	Statistiques descriptives	2
0.3	Histogramme de distribution	3
0.4	Boxplot par zone	3
0.5	Taille ménage vs consommation	4
0.6	Corrélation Spearman	5
0.7	Tests de normalité (Shapiro-Wilk)	5
0.8	Homogénéité des variances (Levene)	6
0.9	Test de Wilcoxon	6
0.10	Taille d'effet (Cohen's d)	7
0.11	Régression multivariée	7
0.12	Conclusion	7
0.13	Recommandations	8
0.14	Résumé	8

```
# Chargement et traitement des données depuis le fichier ZIP
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, haven::read_dta), basename(dta_files))

# Extraction et calculs : taille du ménage, consommation totale, zone
roster <- data_list[["s01_me_SEN2018.dta"]] %>% mutate(id = paste(grappe, menage, sep = "_"))
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)

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)

```

0.1 Contexte

- Étude de la consommation alimentaire individuelle
- Variables : zone (urbaine vs rurale), taille du ménage
- Objectif : détecter des différences et effets significatifs

0.2 Statistiques descriptives

```

desc_zone <- df_indiv %>%
group_by(zone) %>%
summarise(
  n = n(),
  moyenne = mean(food_per_indiv),
  ecart_type = sd(food_per_indiv)
)
knitr::kable(desc_zone, caption = "Statistiques par zone")

```

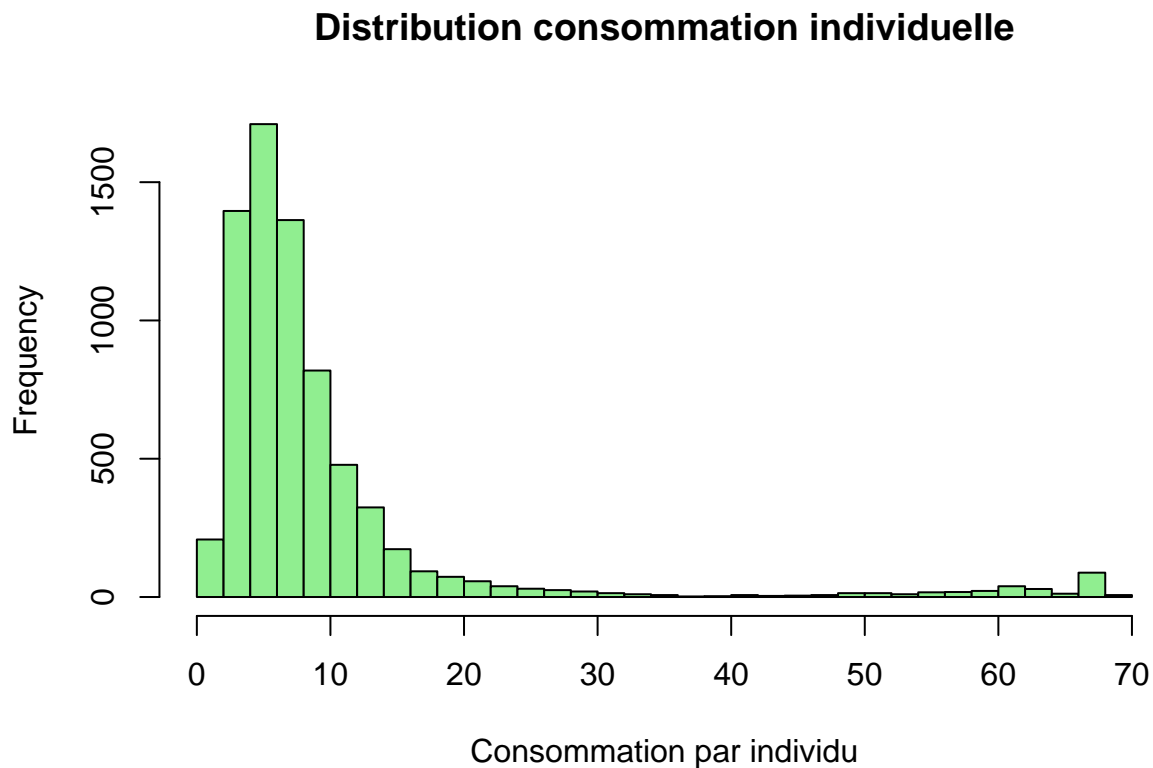
Table 1: Statistiques par zone

zone	n	moyenne	ecart_type
Rurale	3212	7.391587	8.285034
Urbaine	3925	11.384799	13.532003

Interprétation : la moyenne de consommation alimentaire individuelle est plus élevée en zone urbaine qu'en zone rurale, ce qui pourrait refléter des différences de mode de vie ou de pouvoir d'achat.

0.3 Histogramme de distribution

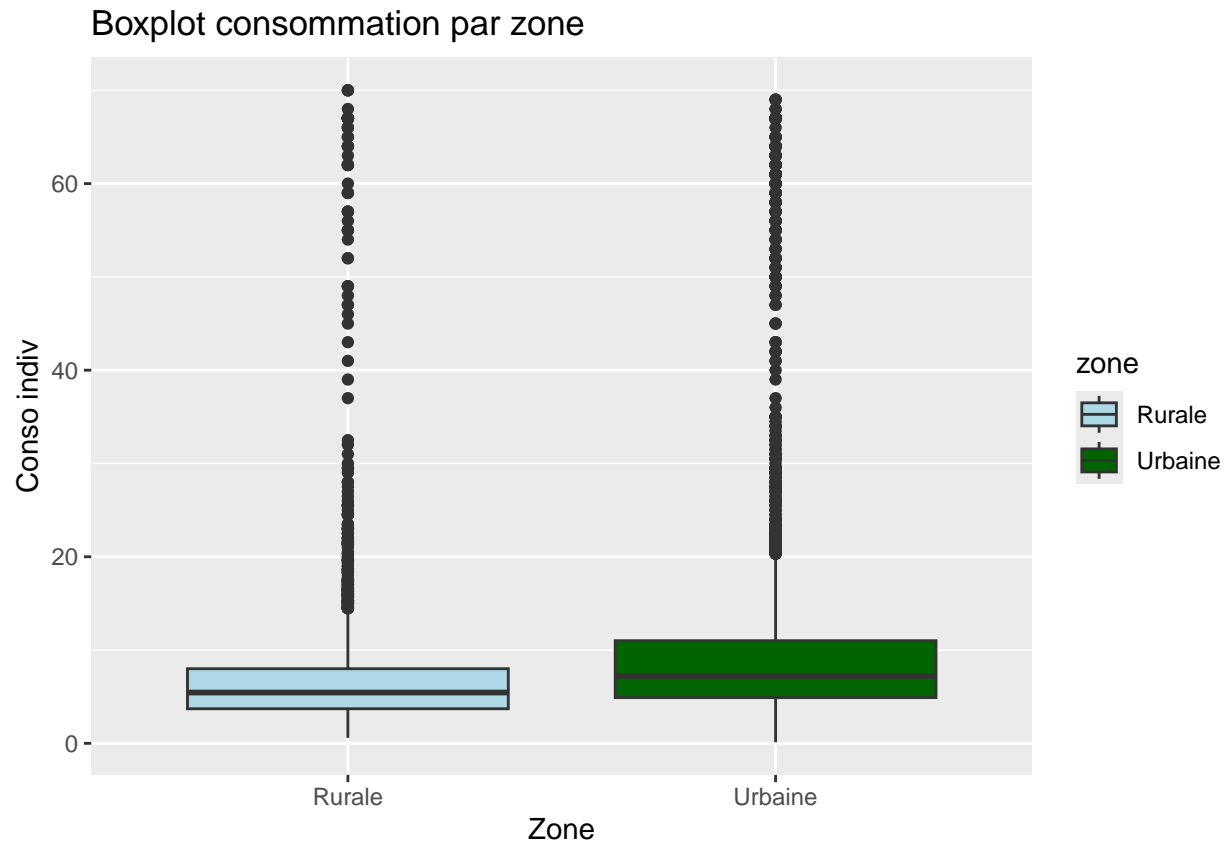
```
hist(df_indiv$food_per_indiv, breaks = 30, col = "lightgreen",  
     main = "Distribution consommation individuelle",  
     xlab = "Consommation par individu")
```



Analyse : La distribution est asymétrique droite avec une longue traîne, ce qui indique une grande hétérogénéité entre ménages.

0.4 Boxplot par zone

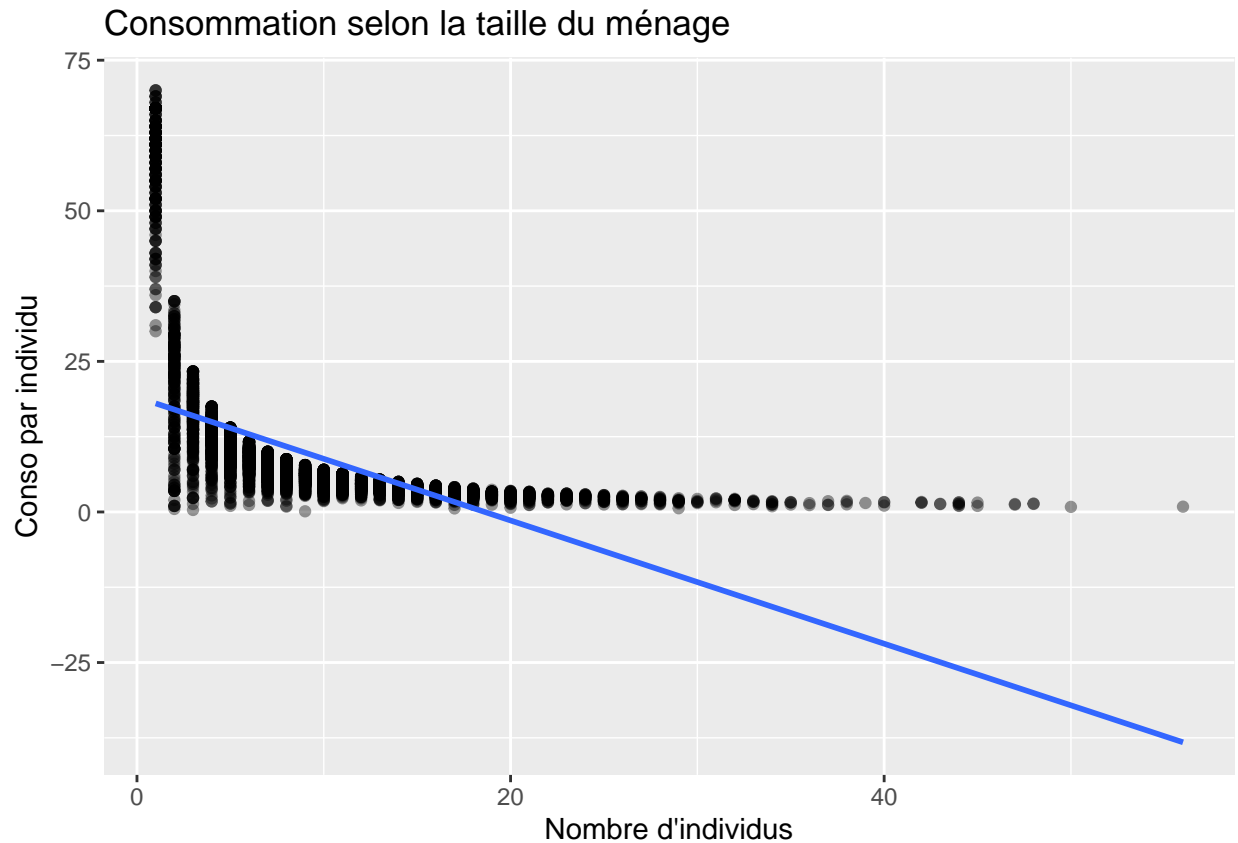
```
ggplot(df_indiv, aes(x = zone, y = food_per_indiv, fill = zone)) +  
  geom_boxplot() +  
  scale_fill_manual(values = c("Urbaine" = "darkgreen", "Rurale" = "lightblue")) +  
  labs(title = "Boxplot consommation par zone", x = "Zone", y = "Conso indiv")
```



Interprétation : En moyenne, les individus urbains consomment plus, avec une distribution plus dispersée.

0.5 Taille ménage vs consommation

```
ggplot(df_indiv, aes(x = n_indiv, y = food_per_indiv)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Consommation selon la taille du ménage",
       x = "Nombre d'individus", y = "Conso par individu")
```



Tendance observée : plus le ménage est grand, moins chaque individu consomme en moyenne.

0.6 Corrélation Spearman

```
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
```

Résultat : Corrélation négative modérée, significative statistiquement ($p < 0.05$).

0.7 Tests 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
```

Conclusion : Les distributions ne sont pas normales ($p < 0.05$), justifiant l'usage de tests non paramétriques.

0.8 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
```

Interprétation : Variances inégales entre zones, ce qui exclut l'usage du test t de Student classique.

0.9 Test de Wilcoxon

```
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
```

Conclusion : Les niveaux de consommation diffèrent significativement entre zones ($p < 0.05$).

0.10 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
```

Magnitude : Faible effet ($d \approx 0.36$), mais statistiquement significatif.

0.11 Régression multivariée

```
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
```

Lecture des coefficients :

- **zoneUrbaine** a un effet positif significatif → les ménages urbains consomment davantage
- **n_indiv** a un effet négatif → effet de dilution par individu
- Le modèle explique plus de 60% de la variance totale ($R^2 = 0.62$)

0.12 Conclusion

- La taille du ménage influence **négativement** la consommation individuelle
- La zone **urbaine** montre une consommation plus élevée
- Le modèle multivarié est significatif ($R^2 > 60\%$)
- Tests non paramétriques utilisés (données non normales)

0.13 Recommandations

- **Ciblage des ménages ruraux** : Considérant leur consommation plus faible, les politiques publiques devraient privilégier ces zones pour renforcer la sécurité alimentaire.
- **Ajustement aux tailles des ménages** : Les programmes de transferts sociaux doivent prendre en compte la composition des ménages pour éviter les effets de sous-estimation de besoin.
- **Attention aux inégalités intra-urbaines** : Une forte variabilité existe également en zone urbaine, ce qui suggère la nécessité d'analyses complémentaires par quintiles ou niveaux de revenu.
- **Utilisation renforcée des modèles multivariés** : Les analyses futures devraient intégrer davantage de variables explicatives (revenus, accès aux marchés, éducation).
- **Perspectives longitudinales** : Un suivi temporel permettrait de mieux comprendre les dynamiques de consommation et les effets des politiques mises en place.

0.14 Résumé

Cette étude a examiné les différences de consommation alimentaire individuelle selon la zone de résidence (urbaine ou rurale) et la taille du ménage, à partir des données de l'enquête **SEN2018**.

- **Statistiquement**, les individus en zone urbaine consomment davantage que ceux en zone rurale.
- **La taille du ménage** joue un rôle important, une augmentation du nombre d'individus réduit la consommation moyenne par personne.
- Les **tests de Shapiro-Wilk** ont confirmé la non-normalité des distributions, justifiant le recours au **test de Wilcoxon**.
- Les **variances sont inégales**, et la **taille d'effet** est faible mais significative.
- Le **modèle de régression multivarié** explique plus de **60 % de la variance**, confirmant l'effet significatif de la zone et du nombre d'individus.

Ces résultats appellent à une **approche ciblée** dans les politiques de sécurité alimentaire, notamment en zones rurales et dans les grands ménages.