Beamer Presentation – Consommation alimentaire par zone

RIRADJIM NGARMOUNDOU Trésor

École Nationale de la Statistique et de l'Analyse Économique de Dakar

15 May 2025

```r
# Chargement 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_p
dta_files <- list.files(out_dir, pattern = "\\.dta$", full.nam
data_list <- setNames(lapply(dta_files, haven::read_dta), base

# Extraction des variables nécessaires
roster <- data_list[["s01_me_SEN2018.dta"]] %>% mutate(id = pa
cons_food <- data_list[["s08b1_me_SEN2018.dta"]] %>% mutate(id
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
  select(id, food_total) %>%
  filter(food_total > 0) %>%
  distinct(id, .keep_all = TRUE)

zone_df <- data_list[["s00_me_SEN2018.dta"]] %>%
```

Contexte

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

Statistiques descriptives

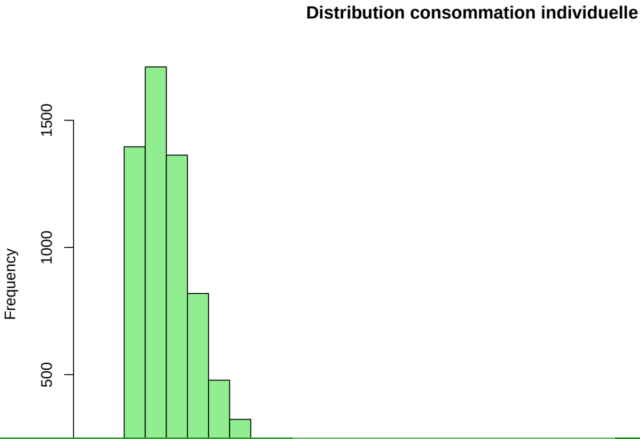
```r
desc_zone <- df_indiv %>%
  group_by(zone) %>%
  summarise(
    n = n(),
    moyenne = mean(food_per_indiv),
    ecart_type = sd(food_per_indiv)
  )
knitr::kable(desc_zone, caption = "Statistiques par zone")
```

Table 1: Statistiques par zone

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

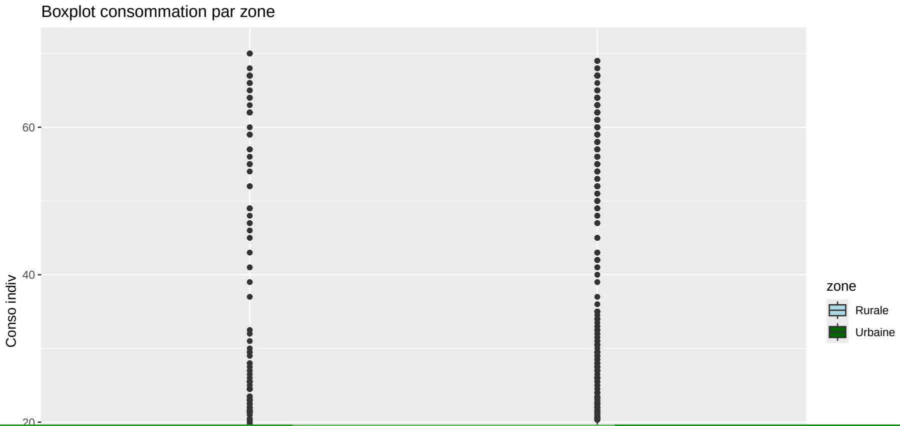
Histogramme de distribution

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



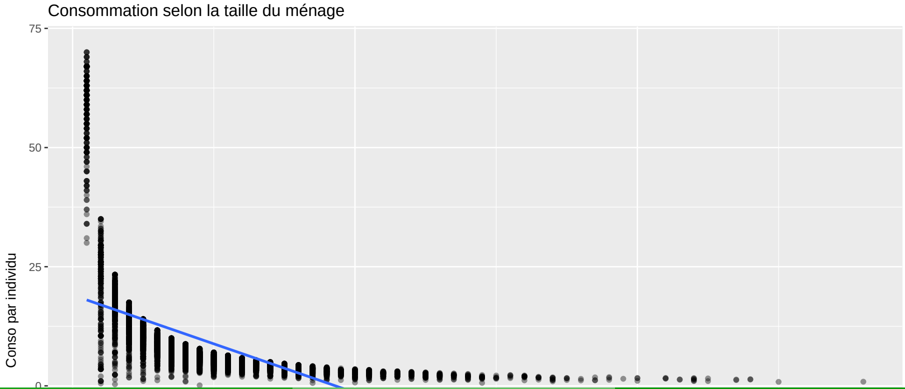
Boxplot par zone

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



Taille ménage vs consommation

```r
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")
```



Corrélation Spearman

```r
cor.test(df_indiv$n_indiv, df_indiv$food_per_indiv, method = '
```

```
##
##  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
```

Tests de normalité (Shapiro-Wilk)

```r
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
```

```r
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
```

Homogénéité des variances (Levene)

```r
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 '
```

Test de Wilcoxon

```r
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
```

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 magnitu
## * <chr>          <chr>  <chr>      <dbl> <int> <int> <ord>
## 1 food_per_indiv Rurale Urbaine   -0.356  3212  3925 small
```

Régression multivariée

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

```
##
## Call:
## lm(formula = log(food_per_indiv + 1) ~ zone + n_indiv, data
##
## 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 ***
## ---
```

Conclusion

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