

# Projet de STA211

*Auteurs :*

Boukary OUEDRAOGO

Bangaly CAMARA

Imad EL HAMMA

*Professeur :*

Mme Niang

2021-2022

$$A_{m,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

$$u''(x_i) \approx \frac{1}{h^2} [u(x_{i+1}) - 2u(x_i) + u(x_{i-1}))]$$

## Introduction

Dans cette partie nous allons explorer les données en réalisant des analyses univariées simples. Nous regarderons les statistiques descriptives simples de chacune des variables, puis nous effectuerons des

*# Analyse univarié simple avec summary*

```
summary(data_menage)
```

```
##  Sexe      Diplome_Max  Type_Prof Occupation Source_ppale_Res Structure_menage
##  1: 549    4          :454   1: 36      1:777      1:845          1: 398
##  2:1018    8          :175   2: 70      2: 80      2: 87          2: 103
##           1          :171   3:435      3: 50      3:493          3:1053
##           9          :168   4:391      4:541      4: 71          4: 13
##           6          :157   5:297      5: 45      5: 71
##           5          :137   6:316      6: 74
##           (Other):305   7: 22

##      Age      Revenus      TYPEMEN      Nb_Enfants_sup_10
##  Min.    :18.00  Min.    : 535  Min.    :1.000  Min.    :0.0000
##  1st Qu.:41.00  1st Qu.:1799  1st Qu.:3.000  1st Qu.:0.0000
##  Median :52.00  Median :2349  Median :5.000  Median :0.0000
##  Mean   :51.87  Mean   :2573  Mean   :4.754  Mean   :0.4359
##  3rd Qu.:61.50  3rd Qu.:2899  3rd Qu.:7.000  3rd Qu.:1.0000
##  Max.    :89.00  Max.    :7600  Max.    :7.000  Max.    :3.0000
##
##  Nb_Enfants_inf_10
##  Min.    :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3165
##  3rd Qu.:0.0000
##  Max.    :3.0000
##
```

```
data_menage %>%
```

```
  describe()
```

```
## Description of .
```

```
##
```

```
## Numeric
```

```
##           mean median      var      sd valid.n
```

```
## Age           51.87     52    209.91   14.49   1567
```

```
## Revenus       2573.20   2349 1967644.23 1402.73   1567
```

```
## TYPEMEN        4.75      5      4.55    2.13   1567
```

```
## Nb_Enfants_sup_10  0.44      0      0.67    0.82   1567
```

```
## Nb_Enfants_inf_10  0.32      0      0.48    0.70   1567
```

```
##
```

```
## Factor
```

```
##
```

```
## Sexe           2      1
```

```
## Count    1018.00 549.00
```

```
## Percent    64.96 35.04
```

```
## Mode 2
```

```
##
```

```
## Diplome_Max     4      8      1      9      6      5      7      3      2
```

```
## Count    454.00 175.00 171.00 168.00 157.00 137.00 132.00 117.00 56.00
```

```
## Percent   28.97  11.17  10.91  10.72  10.02   8.74   8.42   7.47   3.57
```

```
## Mode 4
```

```
##
```

```
## Type_Prof       3      4      6      5      2      1      7
```

```
## Count    435.00 391.00 316.00 297.00 70.00 36.0 22.0
```

```
## Percent   27.76  24.95  20.17  18.95   4.47   2.3   1.4
```

```
## Mode 3
```

```
##
```

```
## Occupation       1      4      2      6      3      5
```

```
## Count    777.00 541.00 80.00 74.00 50.00 45.00
```

```
## Percent   49.59  34.52   5.11   4.72   3.19   2.87
```

```
## Mode 1
```

```
##
```

```
## Source_ppale_Res     1      3      2      4      5
```

```
## Count    845.00 493.00 87.00 71.00 71.00
```

Dans cette partie, on sépare le jeu de données initiales en deux. Une partie contenant les variables catégorielles et une partie contenant les variables numériques.

```
num_cols <- unlist(lapply(data_menage, is.numeric))
cat_col  <- unlist(lapply(data_menage, is.character))
num_data <- data_menage[, num_cols]
cat_data <- data_menage[, cat_col]

data_hist <- num_data %>%
  gather(key="text", value="value") %>%
  mutate(text = gsub("\\\\.", "-", text)) %>%
  mutate(value = round(as.numeric(value), 0))

# p <- data_hist %>%
#   mutate(text = fct_reorder(text, value)) %>%
#   ggplot(aes(x=value, color=text, fill=text)) +
#     geom_histogram(alpha=0.6, binwidth = 5) +
#     # scale_fill_viridis(discrete=TRUE) +
#     # scale_color_viridis(discrete=TRUE) +
#     theme_ipsum() +
#     theme(
#       legend.position="none",
#       panel.spacing = unit(0.1, "lines"),
#       strip.text.x = element_text(size = 8)
#     ) +
#     xlab("") +
#     facet_wrap(~text)
# p
```

Les corrélations entre les variables numériques sont faibles dans la majorité des cas.

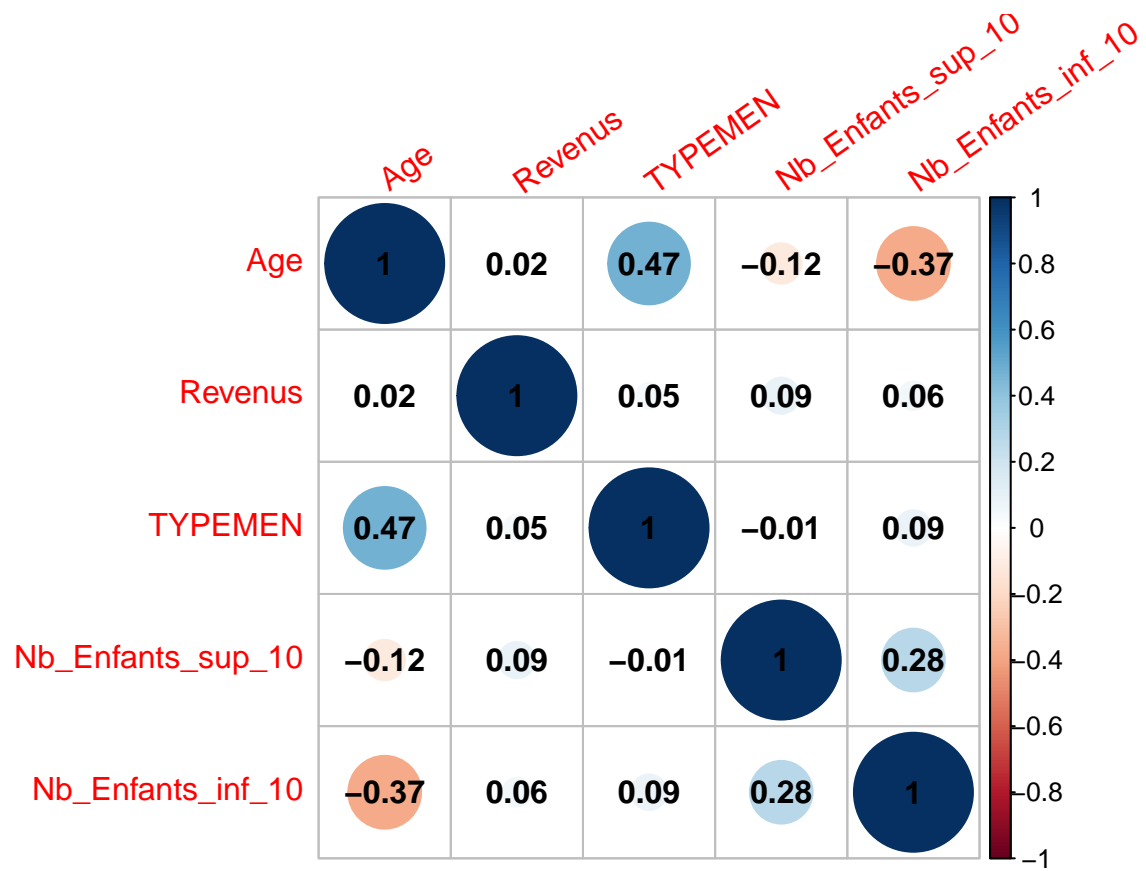


FIGURE 1 – Matrice de corrélation entre les variables numériques du jeu de données