

# Projet d'apprentissage non supervisé

Marie Guibert - Clémence Chesnais

2023-03-23

## Environnement de travail

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stargazer)
```

```
##
## Please cite as:
##
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

## Question 1

### Importation des données

```
pays <- read.csv("Pays_donnees.csv", sep=";", dec=".", stringsAsFactors = T)
# nrow(pays) # On a bien 167 individus
str(pays)
```

```
## 'data.frame':   167 obs. of  10 variables:
## $ pays          : Factor w/ 167 levels "Afghanistan",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ enfant_mort    : num  90.2 16.6 27.3 119 10.3 14.5 18.1 4.8 4.3 39.2 ...
## $ exports        : num  10 28 38.4 62.3 45.5 18.9 20.8 19.8 51.3 54.3 ...
## $ sante          : num  7.58 6.55 4.17 2.85 6.03 8.1 4.4 8.73 11 5.88 ...
## $ imports        : num  44.9 48.6 31.4 42.9 58.9 16 45.3 20.9 47.8 20.7 ...
```

```
## $ revenu      : int  1610 9930 12900 5900 19100 18700 6700 41400 43200 16000 ...
## $ inflation    : num  9.44 4.49 16.1 22.4 1.44 20.9 7.77 1.16 0.873 13.8 ...
## $ esper_vie    : num  56.2 76.3 76.5 60.1 76.8 75.8 73.3 82 80.5 69.1 ...
## $ fert         : num  5.82 1.65 2.89 6.16 2.13 2.37 1.69 1.93 1.44 1.92 ...
## $ pib_h        : int  553 4090 4460 3530 12200 10300 3220 51900 46900 5840 ...
```

```
summary(pays)
```

```
##           pays      enfant_mort      exports      sante
## Afghanistan      : 1   Min.      : 2.60   Min.      : 0.109   Min.      : 1.810
## Albania           : 1   1st Qu.:  8.25   1st Qu.: 23.800   1st Qu.:  4.920
## Algeria           : 1   Median : 19.30   Median : 35.000   Median :  6.320
## Angola            : 1   Mean      : 38.27   Mean      : 41.109   Mean      :  6.816
## Antigua and Barbuda: 1   3rd Qu.: 62.10   3rd Qu.: 51.350   3rd Qu.:  8.600
## Argentina         : 1   Max.      :208.00   Max.      :200.000   Max.      :17.900
## (Other)           :161
##      imports      revenu      inflation      esper_vie
## Min.      : 0.0659   Min.      :  609   Min.      : -4.210   Min.      :32.10
## 1st Qu.: 30.2000   1st Qu.: 3355   1st Qu.:  1.810   1st Qu.:65.30
## Median : 43.3000   Median : 9960   Median :  5.390   Median :73.10
## Mean      : 46.8902   Mean      :17145   Mean      :  7.782   Mean      :70.56
## 3rd Qu.: 58.7500   3rd Qu.:22800   3rd Qu.: 10.750   3rd Qu.:76.80
## Max.      :174.0000   Max.      :125000   Max.      :104.000   Max.      :82.80
##
##      fert      pib_h
## Min.      :1.150   Min.      :  231
## 1st Qu.:1.795   1st Qu.: 1330
## Median :2.410   Median : 4660
## Mean      :2.948   Mean      :12964
## 3rd Qu.:3.880   3rd Qu.:14050
## Max.      :7.490   Max.      :105000
##
```

Dans ce jeu de données, nous pouvons observer 10 variables dont 9 numériques et 1 facteur comprenant les différents pays (individus). Nous avons choisi de transformer la variable pays en facteur pour simplifier nos traitements des données.

Afin de pouvoir analyser ces données, nous allons réaliser des statistiques descriptives de base.

## Statistiques descriptives

```
sum(is.na(pays))
```

```
## [1] 0
```

Le jeu de données ne présente pas de valeur manquante, nous n'avons pas besoin de faire de modification de ce point de vue.

Résumé des données :

```
stargazer(pays,type="text",title="Résumé des données",out="resume_donnees.txt")
```

```
##
## Résumé des données
## =====
## Statistic      N      Mean      St. Dev.   Min      Max
## -----
## enfant_mort 167   38.270    40.329    2.600   208.000
## exports     167   41.109    27.412    0.109   200.000
## sante        167    6.816     2.747    1.810    17.900
## imports     167   46.890    24.210    0.066   174.000
## revenu       167 17,144.690 19,278.070 609    125,000
## inflation    167    7.782    10.571   -4.210   104.000
## esper_vie    167   70.556     8.893    32.100   82.800
## fert         167    2.948     1.514    1.150    7.490
## pib_h        167 12,964.160 18,328.710 231    105,000
## -----
```

```
# par(mfrow=c(1,4))
```