ACP_Caractéristiques_Entreprises_Non_Agricoles

Michel TEVOEDJRE

2025-05-15

Contents

```r
packages <- c("kableExtra", "knitr", "dplyr", "FactoMineR", "factoextra", "haven", "stringr")
installed <- rownames(installed.packages())
to_install <- setdiff(packages, installed)
if (length(to_install) > 0) {
  install.packages(to_install)
}
lapply(packages, library, character.only = TRUE)
```

```
## [[1]]
## [1] "kableExtra" "stats"      "graphics"   "grDevices"  "utils"
## [6] "datasets"   "methods"    "base"
##
## [[2]]
## [1] "knitr"      "kableExtra" "stats"      "graphics"   "grDevices"
## [6] "utils"      "datasets"   "methods"    "base"
##
## [[3]]
##  [1] "dplyr"      "knitr"      "kableExtra" "stats"      "graphics"
##  [6] "grDevices"  "utils"      "datasets"   "methods"    "base"
##
## [[4]]
##  [1] "FactoMineR" "dplyr"      "knitr"      "kableExtra" "stats"
##  [6] "graphics"   "grDevices"  "utils"      "datasets"   "methods"
## [11] "base"
```

```
## 
## [[5]]
##  [1] "factoextra" "ggplot2"    "FactoMineR" "dplyr"      "knitr"     
##  [6] "kableExtra" "stats"      "graphics"   "grDevices"  "utils"     
## [11] "datasets"   "methods"    "base"      
## 
## [[6]]
##  [1] "haven"      "factoextra" "ggplot2"    "FactoMineR" "dplyr"     
##  [6] "knitr"      "kableExtra" "stats"      "graphics"   "grDevices" 
## [11] "utils"      "datasets"   "methods"    "base"      
## 
## [[7]]
##  [1] "stringr"    "haven"      "factoextra" "ggplot2"    "FactoMineR"
##  [6] "dplyr"      "knitr"      "kableExtra" "stats"      "graphics"  
## [11] "grDevices"  "utils"      "datasets"   "methods"    "base"      
```

0.1 Introduction

Ce document détaille les étapes de préparation des données et la réalisation d'une Analyse en Composantes Principales (ACP) sur les données SEN2018.

0.2 Chargement des bibliothèques et des données

```r
library(haven)
library(dplyr)
library(FactoMineR)
library(factoextra)
library(stringr)

senegaldata <- "C:/Users/bmd-tech/Documents/ISEP2/Semestre 2/Traitements Statistiques avec R/SEN2018_me
ma_base <- read_dta(paste0(senegaldata, "s10_2_me_SEN2018.dta"))
```

0.3 Inspection des variables

```r
var_types_labels <- data.frame(
  Variable = names(ma_base),
  Type = sapply(ma_base, function(x) class(x)[1]),
  Label = sapply(ma_base, function(x) attr(x, "label"))
)
kable(head(var_types_labels), caption = "Aperçu des variables et labels") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"), full_width = F)
```

Table 1: Aperçu des variables et labels

	Variable	Type	Label
vague	vague	numeric	Vague
grappe	grappe	numeric	grappe

2

menage	menage	numeric	Identifiant du ménage
s10q12a_1	s10q12a_1	numeric	10.12a.Lister entreprises possédées par le MEN: NUMERO ORDRE ENTREPRISE
s10q12a_2	s10q12a_2	character	10.12a.Lister entreprises possédées par le MEN: NOM ENTREPRISE
s10q13	s10q13	numeric	10.13.Quel est le répondant principal, pour cette entreprise?

0.4 Sélection et nettoyage des variables

```
vars_quant <- ma_base %>%
  select(where(is.numeric)) %>%
  select(where(~ mean(is.na(.)) < 0.3)) %>%
  select(where(~ sd(., na.rm = TRUE) > 0)) %>%
  select(-any_of(c("vague", "grappe", "menage")))
```

0.5 Imputation des valeurs manquantes

```
vars_imputed <- vars_quant %>%
  mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))
```

0.6 Vérification des données

```
stopifnot(!any(is.na(vars_imputed)))
stopifnot(all(sapply(vars_imputed, sd) > 0))
```
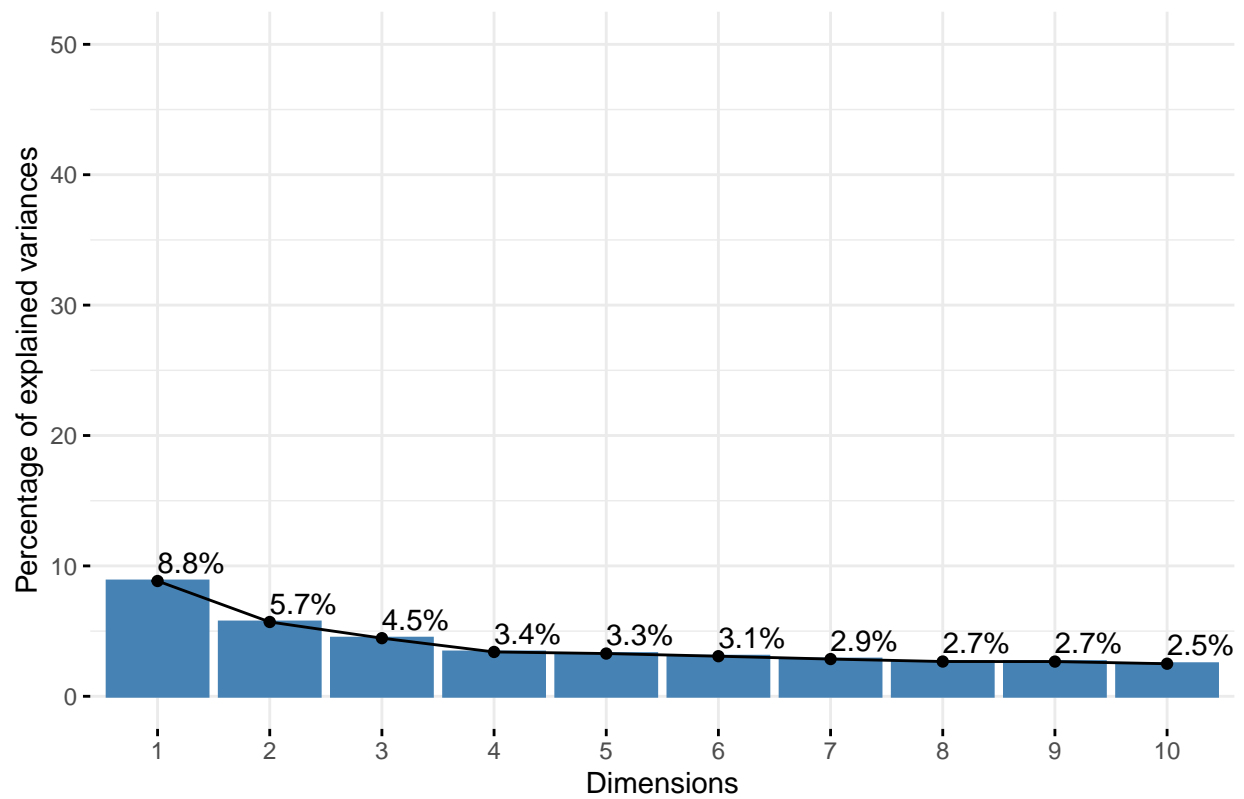
0.7 Standardisation et ACP

```
vars_scaled <- scale(vars_imputed)
res_acp <- PCA(vars_scaled, graph = FALSE)
```
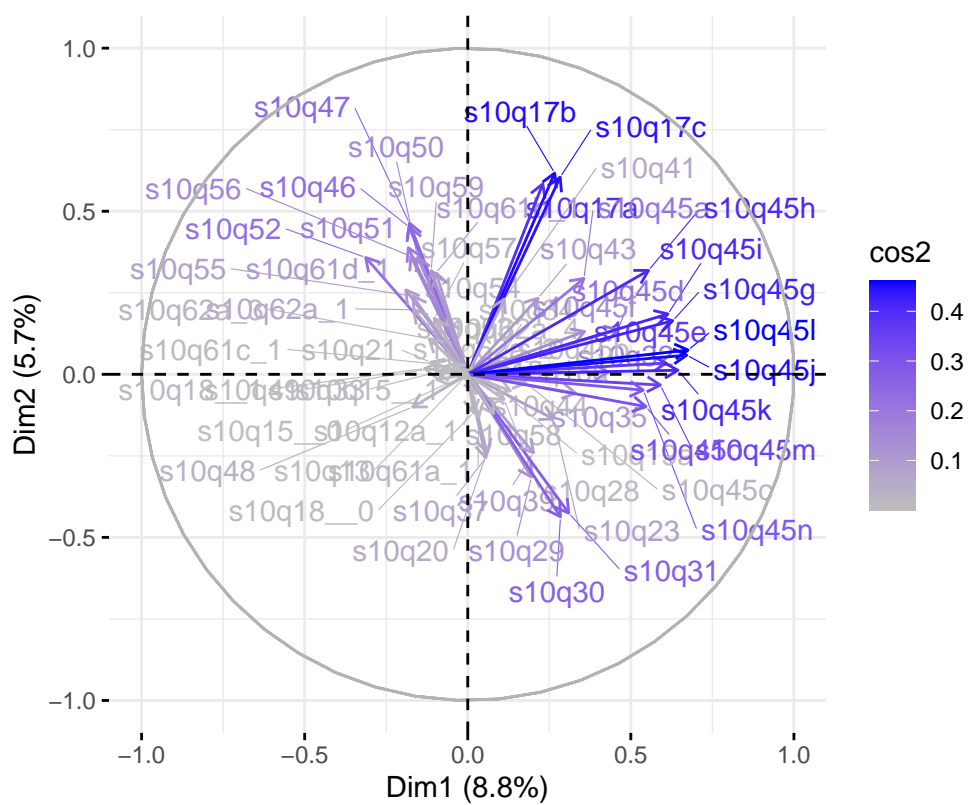
0.8 Visualisation des résultats

```
fviz_eig(res_acp, addlabels = TRUE, ylim = c(0, 50))
```

Scree plot


```
fviz_pca_var(res_acp, col.var = "cos2", gradient.cols = c("grey", "blue"), repel = TRUE)
```

Variables – PCA


```
fviz_pca_ind(res_acp, col.ind = "cos2", gradient.cols = c("grey", "red"), repel = TRUE)
```

Individuals – PCA

0.9 Résultats : Scores et Contributions

```r
coord_individus <- as.data.frame(res_acp$ind$coord)
coord_individus <- cbind(id = rownames(coord_individus), coord_individus)
kable(head(coord_individus), caption = "Coordonnées des individus (ACP)") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"), full_width = F)
```

Table 2: Coordonnées des individus (ACP)

id	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1	2.1118065	-1.154083	2.1262173	0.1618884	-0.4785265
2	-0.0801687	-1.617607	1.4680542	0.4330290	-1.8265737
3	1.4070552	-1.609932	2.0917020	-0.0110245	-0.8015752
4	-4.0224312	-2.236942	-0.1002049	-0.1125851	-0.7713629
5	0.2281184	-1.471436	0.9641094	1.2906176	-1.2594238
6	0.9597069	-1.713737	1.6883380	0.5803354	-1.3317529

```r
contrib_variables <- as.data.frame(res_acp$var$contrib)
contrib_variables <- cbind(variable = rownames(contrib_variables), contrib_variables)
kable(head(contrib_variables), caption = "Contribution des variables aux axes principaux") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"), full_width = F)
```

Table 3: Contribution des variables aux axes principaux

	variable	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
s10q12a_1	s10q12a_1	0.0119412	0.0972148	0.1227384	4.1529441	13.2207577
s10q13	s10q13	0.0000001	0.0000002	0.0000021	0.0003792	0.0021928
s10q15___0	s10q15___0	0.0000002	0.0000010	0.0000056	0.0007003	0.0036191
s10q15___1	s10q15___1	0.1608980	0.0075776	0.0335268	0.0728582	1.4428225
s10q17a	s10q17a	0.9351112	9.3470472	14.7713536	1.4455455	0.0337849
s10q17b	s10q17b	1.2542284	10.4340725	13.5770239	1.3169370	0.0000542