
```
title: "Projet ACP"
author: "Diène Diouf"
output: Html_document
```

Dans ce projet, nous allons effectuer l'analyse d'une base nommée auto. Comme nous le verrons, cette base est un tableau rectangulaire constitué d'individus en lignes et en colonnes de variables quantitatives. La technique la plus indiquée pour analyser ce type de données est l'analyse en composantes principales (ACP). Elle nous permettra de mettre en lumière les similitudes et différences entre les individus mais aussi les liaisons entre les variables. En d'autres termes, il s'agira d'obtenir un résumé pertinent et fidèle des données initiales qui, en gardera donc le maximum d'informations.

Pour les besoins du projet, nous allons utiliser les packages factominer, utilisé pour effectuer des analyses multivariées; factoextra pour la visualisation avancée des données et tidyverse qui, est une compilation de plusieurs autres packages et bien d'autres, utiles à l'analyse.

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(factoextra)
library(FactoMineR)
library(readxl)
```
```

Nous allons importer la base de données .

```
```{r import-data}
auto<-read_excel("C:\\Users\\Asus\\Documents\\site web\\Mes
projets\\auto.xls")
```
```

Dans les lignes qui suivent, nous allons afficher les 5 premières lignes et 5 dernières de la base. Cela nous permettra de nous faire une idée sur la disposition des données : les noms des colonnes et leur nombre et un aperçu sur la nature des lignes.

```
```{r Les cinq premières lignes de la base auto}
```

```
head(auto, n=5)
```

```
```
```

A tibble: 5 x 7

| Modele
<chr> | Cylindree
<dbl> | Puissance
<dbl> | Vitesse
<dbl> | Poids
<dbl> | Largeur
<dbl> | Longueur
<dbl> |
|---------------------|--------------------|--------------------|------------------|----------------|------------------|-------------------|
| Citroen C2 1.1 Base | 1124 | 61 | 158 | 932 | 1659 | 3666 |
| Smart Fortwo Coupe | 698 | 52 | 135 | 730 | 1515 | 2500 |
| Mini 1.6 170 | 1598 | 170 | 218 | 1215 | 1690 | 3625 |
| Nissan Micra 1.2 65 | 1240 | 65 | 154 | 965 | 1660 | 3715 |
| Renault Clio 3.0 V6 | 2946 | 255 | 245 | 1400 | 1810 | 3812 |

```
```{r Les cinq dernières lignes de la base auto}
```

```
tail(auto, n=5)
```

```
```
```

A tibble: 5 x 7

| Modele
<chr> | Cylindree
<dbl> | Puissance
<dbl> | Vitesse
<dbl> | Poids
<dbl> | Largeur
<dbl> | Longueur
<dbl> |
|-------------------------------|--------------------|--------------------|------------------|----------------|------------------|-------------------|
| Renault Scenic 1.9 dCi 120 | 1870 | 120 | 188 | 1430 | 1805 | 4259 |
| Volkswagen Touran 1.9 TDI 105 | 1896 | 105 | 180 | 1498 | 1794 | 4391 |
| Land Rover Defender Td5 | 2495 | 122 | 135 | 1695 | 1790 | 3883 |
| Land Rover Discovery Td5 | 2495 | 138 | 157 | 2175 | 2190 | 4705 |
| Nissan X-Trail 2.2 dCi | 2184 | 136 | 180 | 1520 | 1765 | 4455 |

```
```{r vérifier s'il y a des Valeurs manquantes}
```

```
valeurs_manquantes <- sum(is.na(auto))
```

```
if (valeurs_manquantes > 0) {
```

```
 cat("Le nombre de valeurs manquantes de votre base est",
 valeurs_manquantes)
```

```
} else{
```

```
 ("Il n' y a pas de valeurs manquantes dans la base auto, elle semble
 propre")
```

```
}
```

```
```
```

```
[1] "Il n' y a pas de valeurs manquantes dans la base auto, elle semble propre"
```

...

• • •

```
A tibble: 1 x 7
```

| Modele
<chr> | Cylindree
<dbl> | Puissance
<dbl> | Vitesse
<dbl> | Poids
<dbl> | Largeur
<dbl> | Longueur
<dbl> |
|-----------------|--------------------|--------------------|------------------|----------------|------------------|-------------------|
| Ferrari Enzo | 5998 | 660 | 350 | 1365 | 2650 | 4700 |

```
1 row
```

```

```{r Voiture la plus lente}

voiture_lente <- auto %>% arrange(Vitesse) %>% head(1)

voiture_lente
```

```

A tibble: 1 x 7

| Modele
<chr> | Cylindree
<dbl> | Puissance
<dbl> | Vitesse
<dbl> | Poids
<dbl> | Largeur
<dbl> | Longueur
<dbl> |
|--------------------|--------------------|--------------------|------------------|----------------|------------------|-------------------|
| Smart Fortwo Coupe | 698 | 52 | 135 | 730 | 1515 | 2500 |

1 row

```

```{r Matrice des corrélations}

Sélectionner les variables numériques de la base de données auto
variables_numeriques <- auto[, sapply(auto, is.numeric)]

cor(variables_numeriques)
```

```

| | cylindree | Puissance | Vitesse | Poids |
|-----------|-----------|-----------|-----------|-----------|
| cylindree | 1.0000000 | 0.9540309 | 0.8853884 | 0.6922064 |
| Puissance | 0.9540309 | 1.0000000 | 0.9337772 | 0.5287833 |
| Vitesse | 0.8853884 | 0.9337772 | 1.0000000 | 0.4663310 |
| Poids | 0.6922064 | 0.5287833 | 0.4663310 | 1.0000000 |
| Largeur | 0.7060659 | 0.7299825 | 0.6187895 | 0.4771765 |
| Longueur | 0.6638563 | 0.5268120 | 0.5780593 | 0.7947887 |
| | Largeur | Longueur | | |
| Cylindree | 0.7060659 | 0.6638563 | | |
| Puissance | 0.7299825 | 0.5268120 | | |
| Vitesse | 0.6187895 | 0.5780593 | | |
| Poids | 0.4771765 | 0.7947887 | | |
| Largeur | 1.0000000 | 0.5912533 | | |
| Longueur | 0.5912533 | 1.0000000 | | |

Nous remarquons que les variable cylindree et puissance sont positivement très corrélées (0,954) ce qui suggère une évolution commune. C'est le cas aussi des variables vitesse et puissance (0,934) et vitesse et cylindree (0,885). Grosso modo, toutes les variables sont corrélées positivement. Il existe donc un facteur taille qui permet de les ordonner des plus petites aux plus grandes.

```

```{r Application de la méthode}

colonnes <- c("Cylindree", "Puissance", "Vitesse", "Poids", "Largeur",
"Longueur")

acp_auto <- PCA(auto[, colonnes], graph = FALSE)
```

```

```
```{r}
```

```
summary.PCA(acp_auto)
```

```
```
```

```
Call:
PCA(X = auto[, colonnes], graph = FALSE)

Eigenvalues
```

| | Dim.1 | Dim.2 | Dim.3 |
|----------------------|--------|--------|---------|
| Variance | 4.411 | 0.853 | 0.436 |
| % of var. | 73.521 | 14.223 | 7.261 |
| Cumulative % of var. | 73.521 | 87.745 | 95.006 |
| | Dim.4 | Dim.5 | Dim.6 |
| Variance | 0.236 | 0.051 | 0.012 |
| % of var. | 3.931 | 0.857 | 0.206 |
| Cumulative % of var. | 98.937 | 99.794 | 100.000 |

Les résultats de l'ACP montrent d'abord le tableau des valeurs propres. Dans ce tableau, on peut voir les variances des trois premières composantes principales.

Ainsi, le premier axe factoriel contient 73,52% de l'inertie de la base. Le deuxième axe en contient 14,22%, ce qui fait que les deux premières composantes principales, à elles seules, contiennent 87,74% de l'information. Pour réduire la dimensionnalité, il semble donc intéressant de retenir ces deux axes.

```
Individuals (the 10 first)
```

| | Dist | Dim.1 | ctr | cos2 | Dim.2 |
|----|-------|--------|--------|-------|--------|
| 1 | 2.665 | -2.596 | 6.365 | 0.949 | -0.510 |
| 2 | 4.575 | -4.150 | 16.269 | 0.823 | -1.666 |
| 3 | 1.713 | -1.382 | 1.804 | 0.651 | -0.816 |
| 4 | 2.571 | -2.513 | 5.967 | 0.955 | -0.404 |
| 5 | 1.078 | -0.003 | 0.000 | 0.000 | -0.916 |
| 6 | 1.180 | -1.121 | 1.186 | 0.902 | 0.169 |
| 7 | 1.853 | -1.725 | 2.812 | 0.867 | 0.300 |
| 8 | 0.871 | 0.553 | 0.288 | 0.402 | 0.523 |
| 9 | 0.823 | 0.078 | 0.006 | 0.009 | 0.482 |
| 10 | 1.183 | 0.838 | 0.663 | 0.502 | 0.460 |

| | ctr | cos2 | Dim.3 | ctr | cos2 |
|----|--------|-------|--------|-------|-------|
| 1 | 1.270 | 0.037 | 0.179 | 0.307 | 0.005 |
| 2 | 13.550 | 0.133 | -0.274 | 0.720 | 0.004 |
| 3 | 3.249 | 0.227 | -0.372 | 1.326 | 0.047 |
| 4 | 0.795 | 0.025 | 0.174 | 0.289 | 0.005 |
| 5 | 4.092 | 0.722 | -0.385 | 1.419 | 0.128 |
| 6 | 0.139 | 0.020 | 0.168 | 0.271 | 0.020 |
| 7 | 0.441 | 0.026 | 0.415 | 1.643 | 0.050 |
| 8 | 1.336 | 0.360 | -0.256 | 0.629 | 0.087 |
| 9 | 1.133 | 0.342 | -0.526 | 2.649 | 0.408 |
| 10 | 1.033 | 0.151 | -0.164 | 0.258 | 0.019 |

Le deuxième tableau met en exergue, les coordonnées des 10 premiers individus, leurs contributions à la formation des axes et leurs qualités de représentations.

| Variables | | | | | |
|-----------|-------|--------|--------|--------|--------|
| | Dim.1 | ctr | cos2 | Dim.2 | ctr |
| Cylindree | 0.962 | 20.997 | 0.926 | -0.127 | 1.888 |
| Puissance | 0.923 | 19.326 | 0.853 | -0.353 | 14.573 |
| Vitesse | 0.886 | 17.800 | 0.785 | -0.339 | 13.446 |
| Poids | 0.757 | 12.988 | 0.573 | 0.576 | 38.837 |
| Largeur | 0.801 | 14.553 | 0.642 | -0.111 | 1.444 |
| Longueur | 0.795 | 14.337 | 0.632 | 0.504 | 29.812 |
| | cos2 | Dim.3 | ctr | cos2 | |
| Cylindree | 0.016 | -0.142 | 4.596 | 0.020 | |
| Puissance | 0.124 | -0.090 | 1.874 | 0.008 | |
| Vitesse | 0.115 | -0.206 | 9.772 | 0.043 | |
| Poids | 0.331 | -0.147 | 4.990 | 0.022 | |
| Largeur | 0.012 | 0.583 | 77.964 | 0.340 | |
| Longueur | 0.254 | 0.059 | 0.805 | 0.004 | |

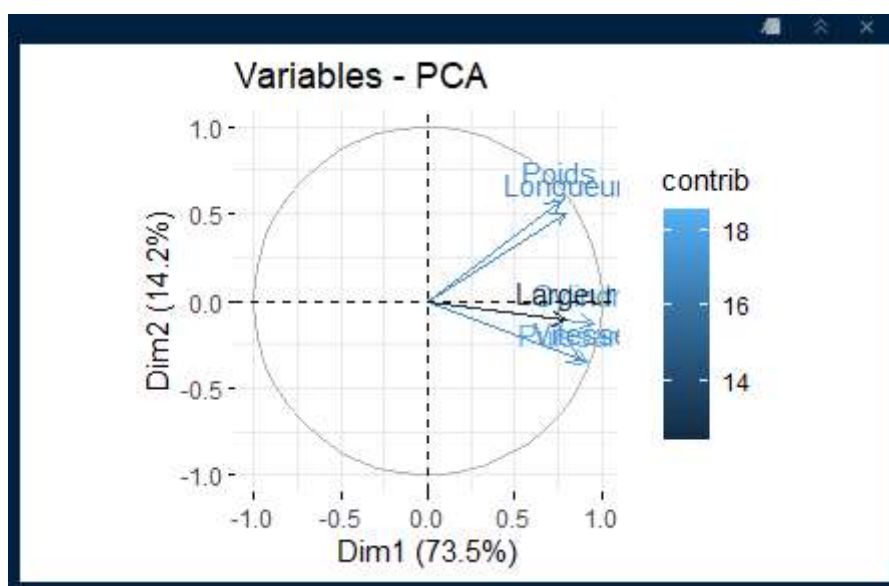
Le troisième tableau concerne les variables.

La variable cylindree par exemple, contribue grandement à la formation de l'axe 1 (20,977), est bien représentée (cos2 proche de 1).

```

```{r}
fviz_pca_var(acp_auto, col.var = "contrib")
```

```



Comme mentionné avant, la lecture du graphique confirme que toutes les variables sont corrélées positivement. Elles sont du même côté sur l'axe 1 et contribuent donc dans le même sens à sa formation.

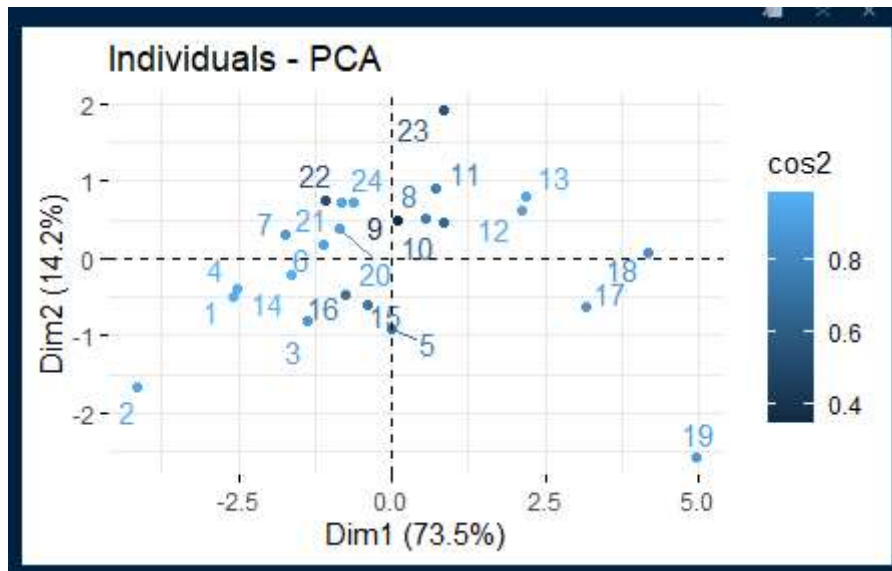
Sur l'axe 2, il y a une opposition entre d'une part les variables poids et longueur (qui sont corrélées positivement) et d'autre part, puissance et vitesse (qui sont également corrélées positivement).

```

```{r}

fviz_pca_ind(acp_auto, col.ind = "cos2",
repel = TRUE)
```

```



La carte factorielle des individus montre les oppositions entre les individus. Les points en bleu clair représentent les véhicules les mieux représentés sur les axes. Ce sont eux qui sont pris en compte dans l'interprétation.

Nous constatons par exemple une opposition sur l'axe 1 entre les voitures Smart Fortwo Coupé (2), Nissan Micra (4), Citroen (1) et les voitures Bentley Continental (18), BMW 745i (12).

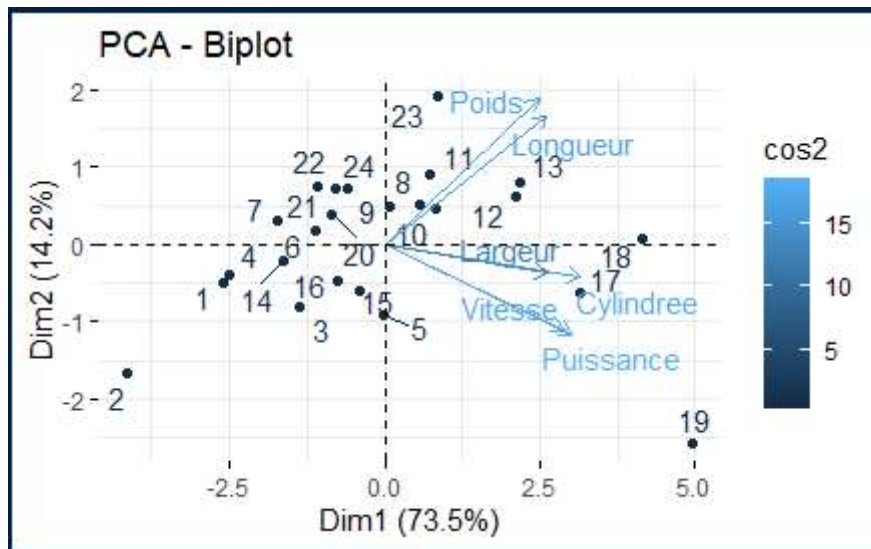
L'axe 2 oppose lui, Volkswagen Touran (21), Land Rover Discovery (23) à Ferrari enzo (19), Renault Clio (5).

```

```{r}

fviz_pca_biplot(acp_auto, col.ind = "cos2", col.ind.sup = "blue", col.var =
"contrib", repel = TRUE)
```

```

Le biplot montre la représentation simultanée des individus et des variables. Il permet de faire le parallèle entre les deux analyses.

Ainsi, nous constatons par exemple une opposition sur l'axe 1 entre les voitures Smart Fortwo Coupé (2), Nissan Micra (4), Citroen(1) qui, sont petites, légères et les voitures Bentley Continental (18), BM 745 (12) qui, sont de grosses voitures qui sont de grosses cylindrées.

L'axe 2 oppose lui, Volkswagen Touran (21), Land Rover Discovery (23) qui sont lourdes, longues et relativement lente à Ferrari enzo (19), Renault Clio (5) qui, sont rapides, puissantes.

Ainsi, l'analyse en composantes principales de cette base a permis de visualiser les oppositions et les similitudes entre les voitures en fonction de caractéristiques comme, la vitesse, la taille, le poids...