

J'ai utilisé R et RStudio pour effectuer les différents calculs. Après avoir transformé le fichier xls en csv, je l'ai importé avec la commande :

```
data <- read.csv2('Geraud.csv', sep=';', header=T)
```

Corrélations

La matrice ci-dessous montre les corrélations entre les variables :

	Var1	Var2	Var3	Var4	Var5	Var6	Var7
Var1	1.000	0.705	0.527	0.016	0.318	0.001	0.281
Var2	0.705	1.000	0.330	0.186	0.384	0.058	0.408
Var3	0.527	0.330	1.000	0.181	-0.417	-0.028	-0.392
Var4	0.016	0.186	0.181	1.000	-0.171	-0.573	-0.073
Var5	0.318	0.384	-0.417	-0.171	1.000	0.017	0.883
Var6	0.001	0.058	-0.028	-0.573	0.017	1.000	-0.031
Var7	0.281	0.408	-0.392	-0.073	0.883	-0.031	1.000

On obtient cette matrice avec la commande suivante :

```
round(cor(data),3)
```

On peut voir quelques variables qui sont fortement corrélées :

- Var1 et Var2, avec un coefficient de 0.705
- Var7 et Var5, avec un coefficient de 0.883

D'autres sont très peu corrélées :

- Var1 et Var4, avec 0.016
- Var1 et Var6, avec 0.001
- Var5 et Var6, avec 0.017
- Var7 et Var4, avec -0.073

ACP

D'après la matrice « Importance of components », on voit que la somme des proportions de variance des 2 premiers facteurs est de 64,3%, ce qui est inférieur à 80%. Il faut donc prendre en compte également le **troisième** facteur pour atteindre une proportion de variance de 85,7%.

Les individus Ind33 et Ind20 sont **proches** car ils sont proches sur le schéma et leurs tailles sont également similaires.

Les individus Ind14 et Ind64 sont **proches** car leurs coordonnées sont similaires :

- Ind14 a pour coordonnées (0.69, 0.05)
- Ind64 a pour coordonnées (0.74, -0.1)

Les individus Ind55 et Ind40 sont placés à l'opposé l'un de l'autre, ils sont donc **éloignés**.

L'axe 1 correspond à la première composante principale, la variable la mieux représentée est **Var5**, car il s'agit de la variable avec la corrélation la plus importante entre les variables d'origine et les composantes principales.

De manière analogue, la variable la moins bien représentée sur l'axe 2 est **Var6**, car la corrélation est la plus faible.

Pour déterminer la variable la mieux représentée sur les deux axes, on calcule la somme des carrés des corrélations sur les deux premiers axes, cela donne alors :

Var1	Var2	Var3	Var4	Var5	Var6	Var7
0.80	0.80	0.78	0.27	0.91	0.09	0.86

On voit donc que la variable **Var6** est la moins bien représentée avec seulement 9% de reconstitution.

Le cercle de corrélation nous indique plusieurs éléments :

- Les variables Var6 et Var4 sont peu représentées : la longueur de la flèche n'atteint pas le bord du cercle
- 3 groupes semblent ensuite se distinguer :
 - Var1 et Var2
 - Var5 et Var7
 - Var3