

TP1-ADM

Guillaume Bernard-Reymond et Lorenzo Gaggini

October 2023

1 Partie 1

Dans cette partie, nous désignerons par (x_i) où $i \in \{1; \dots; 21\}$ les individus et (x^j) où $j \in \{1; \dots; 32\}$ les variables, qu'elles soient quantitatives ou bien qualitatives. Enfin nous noterons (z^j) où $j \in \{4; \dots; 32\}$ les variables quantitatives centrées-réduites.

1. Dans cette question nous placerons dans le cas général où $i \in \{1; \dots; n\}$ et $j \in \{1; \dots; p\}$

- On considère $(w_i)_{i \in \{1 \dots n\}}$ une suite de poids telle que $\sum_{i=1}^n w_i = 1$ et soit $j \in \{1; \dots; p\}$. On peut alors écrire :

$$\begin{aligned} \sum_{i=1}^n w_i z_i^j &= \sum_{i=1}^n w_i \left(\frac{x_i^j - \overline{x^j}}{\sigma_{x^j}} \right) \\ &= \frac{1}{\sigma_{x^j}} \sum_{i=1}^n (w_i x_i^j - w_i \overline{x^j}) \\ &= \frac{1}{\sigma_{x^j}} \left(\overline{x^j} - \overline{x^j} \sum_{i=1}^n w_i \right) \\ &= 0 \end{aligned}$$

Le barycentre du nuage est donc bien $0_{\mathbb{R}^p}$.

Informatiquement, nous avons utiliser les commandes suivantes :

```
Barycentre=(colMeans(CR[4:32]))  
print(Barycentre)
```

Les résultats obtenus étant de l'ordre de 10^{-16} , on peut considérer qu'ils sont égaux à 0.

•

$$\begin{aligned}
 In_O(\{z_i; w_i\}_{i=1, \dots, n}) &= \sum_{i=1}^n w_i \|z_i\|^2 \\
 &= \sum_{i=1}^n w_i \left(\sum_{j=1}^p z_i^j{}^2 \right) \\
 &= \sum_{i=1}^n \left(\sum_{j=1}^p w_i z_i^j{}^2 \right) \\
 &= \sum_{j=1}^p \left(\sum_{i=1}^n w_i z_i^j{}^2 \right)
 \end{aligned}$$

Or l'expression $\sum_{i=1}^n w_i z_i^j{}^2$ n'est rien d'autre que l'expression de la variance de notre variable quantitative centrée réduite qui vaut donc 1.

Ainsi : $In_O(\{z_i; w_i\}_{i=1, \dots, n}) = p$ c'est à dire le nombre de variables quantitatives.

Informatiquement, nous avons utilisé les commandes :

```
#variance
Variance=diag(var(CR[4:32]))
print(Variance)
```

```
#inertia
Inertie=sum(Variance)
print(Inertie)
```

Le résultat obtenu est 29 comme attendu.

2. Pour chaque appellation, nous avons effectué un tri afin d'obtenir pour chaque appellation un tableau des individus. A partir de là, nous avons calculé le poids, les barycentres et les normes euclidiennes carrées de ces trois barycentres. Voici le code utilisé où CR désigne le tableau dont les variables quantitatives sont centrées et réduites :

```
"Q2"
chinon=CR[CR$Label == 'Chinon',]
saumur=CR[CR$Label == 'Saumur',]
bourgueil=CR[CR$Label == 'Bourgueil',]
```

```
#poids
pchi=nrow(chinon)/nrow(CR)
psau=nrow(saumur)/nrow(CR)
```

```
pbou=nrow(bourgueil)/nrow(CR)
```

```
#barycentre
```

```
mchi= unnname(colMeans(chinon[4:32]))
```

```
msau= unnname(colMeans(saumur[4:32]))
```

```
mbou= unnname(colMeans(bourgueil[4:32]))
```

```
#carree des normes euclidiennes des barycentres
```

```
nchi=sum(mchi^2)
```

```
nsau=sum(msau^2)
```

```
nbou=sum(mbou^2)
```

Voici les résultats obtenus :

•	Poids	Norme
Chinon	0,1905	5,1428
Saumur	0,5338	2,0188
Bourgueil	0,2857	3,4319