

TD2 : Analyse des données

Le tableau suivant représente le type d'études poursuivies (université, classes préparatoires, autres) en fonction du parcours suivi au lycée (Lettres, Economie, Maths-Sciences, Technique).

	Univ	Prepa	Autres
Lettres	13.00	2.00	5.00
Economie	20.00	2.00	8.00
Math-Sciences	10.00	5.00	5.00
Tech	7.00	1.00	22.00

```
tab <- matrix(c(13, 2, 5, 20, 2, 8, 10, 5, 5, 7, 1, 22), ncol = 3, byrow = TRUE)
colnames(tab) <- c("Univ", "Prepa", "Autres")
rownames(tab) <- c("Lettres", "Economie", "Math-Sciences", "Tech")
# F: tableau des fréquences
(F <- tab/sum(tab))

##           Univ Prepa  Autres
## Lettres    0.13  0.02   0.05
## Economie   0.20  0.02   0.08
## Math-Sciences 0.10  0.05   0.05
## Tech       0.07  0.01   0.22
```

1) Calculer la matrice X des profils-ligne et la matrice Y des profils-colonne.

```
n <- sum(tab)
ni. <- rowSums(tab)
n.j <- colSums(tab)
D1 <- diag(1/ni.)
(X <- D1 %*% tab)

##           Univ  Prepa  Autres
## [1,] 0.6500 0.10000 0.2500
## [2,] 0.6667 0.06667 0.2667
## [3,] 0.5000 0.25000 0.2500
## [4,] 0.2333 0.03333 0.7333

fi. <- ni./n
diag(1/fi.) %*% F

##           Univ  Prepa  Autres
## [1,] 0.6500 0.10000 0.2500
## [2,] 0.6667 0.06667 0.2667
## [3,] 0.5000 0.25000 0.2500
## [4,] 0.2333 0.03333 0.7333

D2 <- diag(1/n.j)
(Y <- tab %*% D2)

##           [,1] [,2] [,3]
## Lettres    0.26  0.2  0.125
## Economie   0.40  0.2  0.200
## Math-Sciences 0.20  0.5  0.125
## Tech       0.14  0.1  0.550
```

2) Déterminer les axes principaux ainsi que les coordonnées principales (appliquée sur les profils lignes).

```
(SS <- t(tab) %*% D1 %*% tab %*% D2)

##          [,1] [,2] [,3]
## Univ    0.5683 0.5367 0.40542
## Prepa   0.1073 0.1617 0.07542
## Autres  0.3243 0.3017 0.51917

ei.li <- eigen(SS)
(lamda.l <- ei.li$values)

## [1] 1.00000 0.19981 0.04936

(ap.li <- ei.li$vectors)

##          [,1] [,2] [,3]
## [1,] -0.7715 -0.6063 0.7523
## [2,] -0.1543 -0.1704 -0.6509
## [3,] -0.6172 0.7767 -0.1014

(coo.li <- n * D1 %*% tab %*% D2 %*% ap.li)

##          [,1] [,2] [,3]
## [1,] -1.543 -0.4732 0.26374
## [2,] -1.543 -0.4042 0.50156
## [3,] -1.543 -0.5469 -0.93835
## [4,] -1.543 1.0843 -0.05182
```

3) Idem pour les profils colonnes.

```
(QQ <- tab %*% D2 %*% t(tab) %*% D1)

##          [,1] [,2] [,3] [,4]
## Lettres    0.2202 0.2200 0.2112 0.1590
## Economie   0.3300 0.3333 0.3000 0.2467
## Math-Sciences 0.2112 0.2000 0.2563 0.1550
## Tech       0.2385 0.2467 0.2325 0.4393

ei.co <- eigen(QQ)
(ap.co <- ei.co$vectors)

##          [,1] [,2] [,3] [,4]
## [1,] -0.3922 -0.2516 -0.2138 -0.75388
## [2,] -0.5883 -0.3224 -0.6098 0.64302
## [3,] -0.3922 -0.2908 0.7606 0.13304
## [4,] -0.5883 0.8649 0.0630 -0.02217

(lamda.c <- ei.co$values)

## [1] 1.000e+00 1.998e-01 4.936e-02 -5.930e-17

(coo.co <- n * D2 %*% t(tab) %*% D1 %*% ap.co)

##          [,1] [,2] [,3] [,4]
## [1,] -1.961 -0.6443 -0.30100 -6.099e-17
## [2,] -1.961 -0.9054 1.30212 -1.805e-15
## [3,] -1.961 1.0317 0.05072 -8.674e-17
```