

TP AFC

MATHIEU SROUR AND OMAR ALLOUCH

2023-11-15

AFC

1)

```
library("readxl")
#Read data from excel
X <- read_excel("TP_AFC_majeur1718_travail.xlsx", na = " ")
X <- as.data.frame(X[, 2:3])
X <- na.omit(X)
print(X)
```

Importation et visualisation des données X)

##	Sexe	Fonction
## 1	0	4
## 2	1	3
## 3	2	3
## 4	1	3
## 5	1	2
## 6	1	3
## 7	1	5
## 8	2	2
## 9	1	4
## 10	1	2
## 11	1	3
## 12	1	6
## 13	1	4
## 14	1	2
## 15	2	1
## 16	2	6
## 17	2	1
## 18	1	3
## 19	2	6
## 20	2	1
## 21	2	1
## 22	1	1
## 23	1	1
## 24	1	1
## 25	1	3

## 26	1	3
## 27	1	2
## 28	0	2
## 29	2	1
## 30	2	1
## 31	1	3
## 32	2	1
## 33	1	2
## 34	1	3
## 35	1	3
## 36	1	1
## 37	1	5
## 38	1	3
## 39	2	4
## 40	2	1
## 41	2	1
## 42	1	3
## 43	2	1
## 44	1	2
## 45	2	1
## 46	1	2
## 47	1	2
## 48	1	3
## 49	2	4
## 50	2	1
## 51	1	1
## 52	1	4
## 53	2	3
## 54	1	1
## 55	2	1
## 56	2	1
## 57	2	2
## 58	2	2
## 59	0	3
## 60	2	1
## 61	1	2
## 62	2	1
## 63	1	3
## 64	1	3
## 65	2	1
## 66	2	1
## 67	2	3
## 68	2	1
## 69	2	1
## 70	1	2
## 71	1	1
## 72	1	4
## 73	2	1
## 74	2	1
## 75	1	4
## 76	1	3
## 77	0	3
## 78	0	0
## 79	1	1

## 80	2	1
## 81	0	0
## 82	1	5
## 83	1	1
## 84	2	1
## 85	1	3
## 86	1	2
## 87	0	3
## 88	2	4
## 89	1	7
## 90	1	4
## 91	2	1
## 92	1	3
## 93	0	3
## 94	2	3
## 95	1	6
## 96	2	6
## 97	1	6
## 98	1	2
## 99	2	3
## 100	2	3
## 101	2	7
## 102	2	1
## 103	2	1
## 104	2	1
## 105	1	2
## 106	2	4
## 107	1	3
## 108	1	3
## 109	1	2
## 110	1	4
## 111	2	2
## 112	1	3
## 113	1	4
## 114	1	3
## 115	1	3
## 116	0	2
## 117	1	5
## 118	0	0
## 119	1	2
## 120	0	7
## 121	1	4
## 122	0	4
## 123	2	1
## 124	0	7
## 125	1	3
## 126	2	1
## 127	2	7
## 128	2	2
## 129	1	1
## 130	2	1
## 131	1	3
## 132	1	4
## 133	0	7

## 134	1	3
## 135	1	3
## 136	2	1
## 137	1	5
## 138	2	1
## 139	2	1
## 140	1	3
## 141	1	3
## 142	1	6
## 143	1	6
## 144	2	3
## 145	2	2
## 146	2	3
## 147	2	1
## 148	1	3
## 149	2	1
## 150	1	3
## 151	1	1
## 152	1	2
## 153	1	2
## 154	2	2
## 155	1	6
## 156	1	7
## 157	0	0
## 158	2	5
## 159	2	1
## 160	2	1
## 161	2	3
## 162	1	3
## 163	2	1
## 164	2	1
## 165	2	1
## 166	1	4
## 167	0	3
## 168	2	1
## 169	2	4
## 170	1	1
## 171	1	3
## 172	2	1
## 173	1	3
## 174	1	3
## 175	2	1
## 176	2	1
## 177	0	7
## 178	2	1
## 179	2	1
## 180	1	3
## 181	1	2
## 182	1	4
## 183	2	1
## 184	1	6
## 185	1	3
## 186	1	7
## 187	0	0

## 188	0	2
## 189	2	6
## 190	2	3
## 191	2	1
## 192	1	6
## 193	0	0
## 194	1	3
## 195	2	1
## 196	2	6
## 197	1	3
## 198	1	3
## 199	2	1
## 200	0	4
## 201	1	6
## 202	0	0
## 203	1	3
## 204	1	6
## 205	2	3
## 206	1	3
## 207	1	3
## 208	1	2
## 209	1	5
## 210	1	2
## 211	2	1
## 212	1	3
## 213	2	3
## 214	2	1
## 215	2	4
## 216	2	1
## 217	1	3
## 218	2	3
## 219	1	2
## 220	2	3
## 221	2	3
## 222	1	2
## 223	1	6
## 224	2	4
## 225	0	3
## 226	1	1
## 227	1	3
## 228	2	3
## 229	2	7
## 230	1	2
## 231	1	1
## 232	1	2
## 233	2	1
## 234	2	1
## 235	0	0
## 236	2	2
## 237	1	3
## 238	2	2
## 239	1	7
## 240	1	3
## 241	1	6

```
## 242    0    3
## 243    2    1
## 244    2    3
## 245    1    4
## 246    1    1
## 247    2    3
## 248    2    6
```

```
V0 <- with(X, table(Sexe, Fonction))
rownames(V0) <- c("Non répondu", "H", "F")
colnames(V0) <- c(
  "Non répondu", "Administratif", "Technicien (OS)",
  "Ingénieur", "Technicien supérieur", "Direction",
  "Contractuel S1", "Contractuel S2"
)
```

Add column names)

```
k <- 248
V <- V0 / k
print(V)
```

Tableau des fréquences)

```
##          Fonction
## Sexe      Non répondu Administratif Technicien (OS)  Ingénieur
## Non répondu 0.032258065  0.000000000      0.012096774 0.028225806
## H           0.000000000  0.060483871      0.096774194 0.193548387
## F           0.000000000  0.229838710      0.036290323 0.072580645
##          Fonction
## Sexe      Technicien supérieur  Direction Contractuel S1 Contractuel S2
## Non répondu      0.012096774 0.000000000      0.000000000 0.016129032
## H                0.052419355 0.024193548      0.048387097 0.016129032
## F                0.028225806 0.004032258      0.024193548 0.012096774
```

Création des Matrice diagonales des marges en lignes et en colonnes) D_n Matrice diagonale des marges en ligne $f_{i.}$

D_p matrice diagonale des marges en colonne $f_{.j}$

```
##          Non répondu      H      F
## Non répondu 0.1008065 0.0000000 0.0000000
## H           0.0000000 0.4919355 0.0000000
## F           0.0000000 0.0000000 0.4072581
```

```
## [1] "Dp"
```

```
## Non répondu Administratif Technicien (OS) Ingénieur
## Non répondu 0.03225806 0.0000000 0.0000000 0.0000000
## Administratif 0.00000000 0.2903226 0.0000000 0.0000000
## Technicien (OS) 0.00000000 0.0000000 0.1451613 0.0000000
## Ingénieur 0.00000000 0.0000000 0.0000000 0.2943548
## Technicien supérieur 0.00000000 0.0000000 0.0000000 0.0000000
## Direction 0.00000000 0.0000000 0.0000000 0.0000000
## Contractuel S1 0.00000000 0.0000000 0.0000000 0.0000000
## Contractuel S2 0.00000000 0.0000000 0.0000000 0.0000000
## Technicien supérieur Direction Contractuel S1
## Non répondu 0.00000000 0.00000000 0.00000000
## Administratif 0.00000000 0.00000000 0.00000000
## Technicien (OS) 0.00000000 0.00000000 0.00000000
## Ingénieur 0.00000000 0.00000000 0.00000000
## Technicien supérieur 0.09274194 0.00000000 0.00000000
## Direction 0.00000000 0.02822581 0.00000000
## Contractuel S1 0.00000000 0.00000000 0.07258065
## Contractuel S2 0.00000000 0.00000000 0.00000000
## Contractuel S2
## Non répondu 0.00000000
## Administratif 0.00000000
## Technicien (OS) 0.00000000
## Ingénieur 0.00000000
## Technicien supérieur 0.00000000
## Direction 0.00000000
## Contractuel S1 0.00000000
## Contractuel S2 0.04435484
```

####Profils lignes et colonnes)

Profil ligne : répartition de la variable Fonction en fonction de la modalité i de la variable Sexe

```
line_profiles <- solve(Dn) %*% V
print(line_profiles)
```

```
## Fonction
## Non répondu Administratif Technicien (OS) Ingénieur
## Non répondu 0.32 0.0000000 0.12000000 0.2800000
## H 0.00 0.1229508 0.19672131 0.3934426
## F 0.00 0.5643564 0.08910891 0.1782178
## Fonction
## Technicien supérieur Direction Contractuel S1 Contractuel S2
## Non répondu 0.12000000 0.00000000 0.00000000 0.16000000
## H 0.10655738 0.04918033 0.09836066 0.03278689
## F 0.06930693 0.00990099 0.05940594 0.02970297
```

Profil colonne : répartition de la variable Sexe en fonction de la modalité i de la variable Fonction

```
column_profiles <- solve(Dp) %*% t(V)
print(column_profiles)
```

```
## Sexe
## Non répondu H F
```

```
## Non répondu 1.00000000 0.0000000 0.0000000
## Administratif 0.00000000 0.2083333 0.7916667
## Technicien (OS) 0.08333333 0.6666667 0.2500000
## Ingénieur 0.09589041 0.6575342 0.2465753
## Technicien supérieur 0.13043478 0.5652174 0.3043478
## Direction 0.00000000 0.8571429 0.1428571
## Contractuel S1 0.00000000 0.6666667 0.3333333
## Contractuel S2 0.36363636 0.3636364 0.2727273
```

```
S <- t(V) %*% line_profiles %*% solve(Dp)
A <- sqrt(solve(Dp)) %*% t(V) %*% line_profiles %*% sqrt(solve(Dp))
print(A)
```

Création de la matrice diagonalisable)

```
##
## Non répondu Administratif Technicien (OS) Ingénieur
## Non répondu 0.32000000 0.00000000 0.05656854 0.09269186
## Administratif 0.00000000 0.47239693 0.15772464 0.22152327
## Technicien (OS) 0.05656854 0.15772464 0.16342477 0.23186989
## Ingénieur 0.09269186 0.22152327 0.23186989 0.32949543
## Technicien supérieur 0.07077214 0.13635569 0.12306310 0.17576993
## Direction 0.00000000 0.05799842 0.07996700 0.11231315
## Contractuel S1 0.00000000 0.13504301 0.11373849 0.15974500
## Contractuel S2 0.13644846 0.07763607 0.07709705 0.11392861
##
## Technicien supérieur Direction Contractuel S1
## Non répondu 0.07077214 0.00000000 0.00000000
## Administratif 0.13635569 0.05799842 0.13504301
## Technicien (OS) 0.12306310 0.07996700 0.11373849
## Ingénieur 0.17576993 0.11231315 0.15974500
## Technicien supérieur 0.09697367 0.05584954 0.08328161
## Direction 0.05584954 0.04356899 0.05786828
## Contractuel S1 0.08328161 0.05786828 0.08537575
## Contractuel S2 0.07004603 0.02580344 0.04062614
##
## Contractuel S2
## Non répondu 0.13644846
## Administratif 0.07763607
## Technicien (OS) 0.07709705
## Ingénieur 0.11392861
## Technicien supérieur 0.07004603
## Direction 0.02580344
## Contractuel S1 0.04062614
## Contractuel S2 0.07820513
```

2)


```

decomp1 <- eigen(A)
eigen_values1 <- decomp1$values
eigen_vectors1 <- decomp1$vectors

print(eigen_values1)

```

Décomposition spectrale pour déterminer les valeurs et les vecteurs propres de S)

```

## [1] 1.000000e+00 3.859159e-01 2.035248e-01 7.021471e-17 1.876014e-17
## [6] -8.321772e-18 -1.199203e-17 -2.018459e-17

```

```

print(eigen_vectors1)

```

```

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] -0.1796053  0.81713046  0.38434589  0.39027936  0.000000000  0.00000000
## [2,] -0.5388159 -0.47602650  0.68185977  0.07720510  0.007030765 -0.08053810
## [3,] -0.3810004  0.03568836 -0.29550099  0.04095235  0.720957636  0.10685923
## [4,] -0.5425448  0.08393718 -0.39912495 -0.03236017  0.003715730 -0.44349922
## [5,] -0.3045356  0.07569516 -0.09963778 -0.20050677 -0.413714109 -0.01934346
## [6,] -0.1680054 -0.02826703 -0.27179448  0.24952969 -0.517757282 -0.17637093
## [7,] -0.2694080 -0.10189621 -0.20779802  0.29399894 -0.182714415  0.82621355
## [8,] -0.2106059  0.28366503  0.11723473 -0.80628389 -0.086835046  0.26700833
##           [,7]      [,8]
## [1,] 0.00000000  0.00000000
## [2,] 0.06615795  0.03563878
## [3,] 0.37143484 -0.30887614
## [4,] -0.42165938  0.40466011
## [5,] -0.22731972 -0.79258430
## [6,] 0.72509232  0.09949716
## [7,] -0.19937683  0.17792305
## [8,] 0.25035656  0.26425261

```

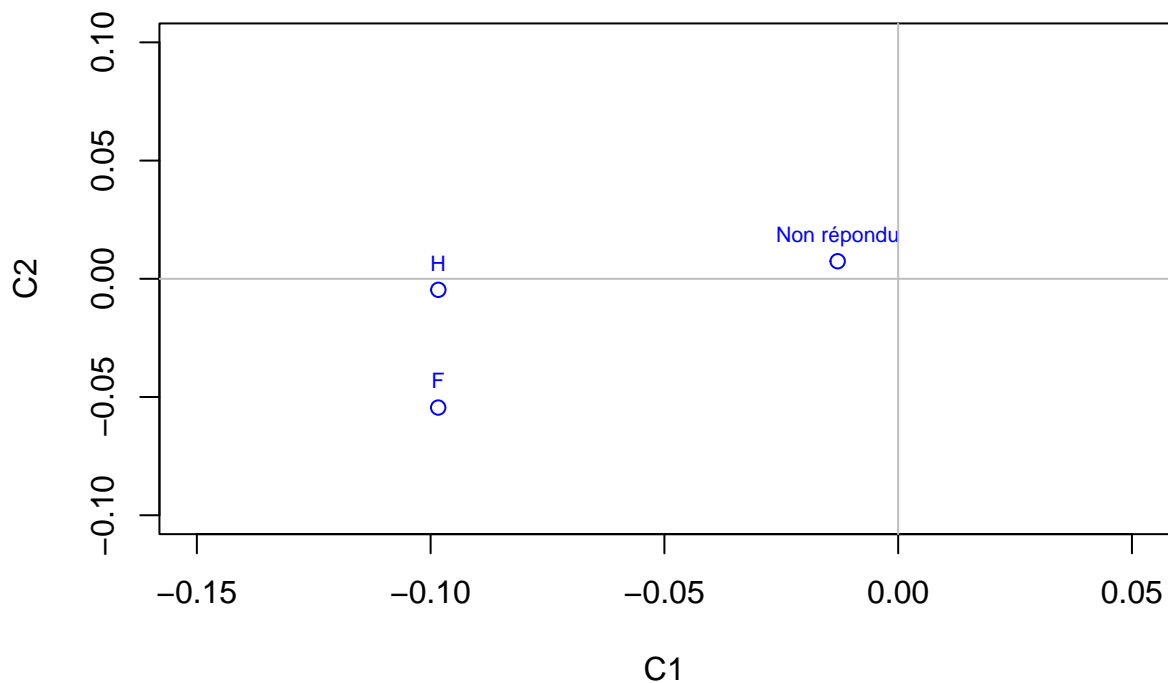
```

# calcul des directions des axes factoriels
u1 <- sqrt(Dp) %*% eigen_vectors1[, 1]
u2 <- sqrt(Dp) %*% eigen_vectors1[, 2]

# nouvelles coordonnées des individus sur le nouveau plan factoriel
C1 <- V %*% u1
C2 <- V %*% u2

# Visualisation du nuage de points dans l'hyperplan obtenu
plot(C1, C2, type = "p", xlim = c(-0.15, 0.05), ylim = c(-0.1, 0.1), col = "blue")
text(C1, C2, rownames(V), cex = 0.7, pos = 3, col = "blue")
abline(h = 0, col = "gray")
abline(v = 0, col = "gray")

```



Matrice à diagonaliser)

```
T <- V %*% column_profiles %*% solve(Dn)

A2 <- sqrt(solve(Dn)) %*% V %*% column_profiles %*% sqrt(solve(Dn))
print(A2)
```

```
##
##           Non répondu           H           F
## Non répondu  0.43068331 0.1765979 0.08915493
## H            0.17659789 0.5953432 0.35687912
## F            0.08915493 0.3568791 0.56341416
```

```
decomp2 <- eigen(A2)
eigen_values2 <- decomp2$values
eigen_vectors2 <- decomp2$vectors

print(eigen_values2)
```

Décomposition spectrale pour déterminer les valeurs et les vecteurs propres de S)

```
## [1] 1.0000000 0.3859159 0.2035248
```

```
print(eigen_vectors2)
```

```
##           [,1]      [,2]      [,3]
## [1,] -0.3175003  0.897351827  0.3065179
## [2,] -0.7013811 -0.004693404 -0.7127710
## [3,] -0.6381677 -0.441290914  0.6308758
```

```
# calcul des directions des axes factoriels
```

```
u1_2 <- sqrt(Dn) %*% eigen_vectors2[, 1]
```

```
u2_2 <- sqrt(Dn) %*% eigen_vectors2[, 2]
```

```
# nouvelles coordonnées des individus sur le nouveau plan factoriel
```

```
C1_2 <- t(V) %*% u1_2
```

```
C2_2 <- t(V) %*% u2_2
```

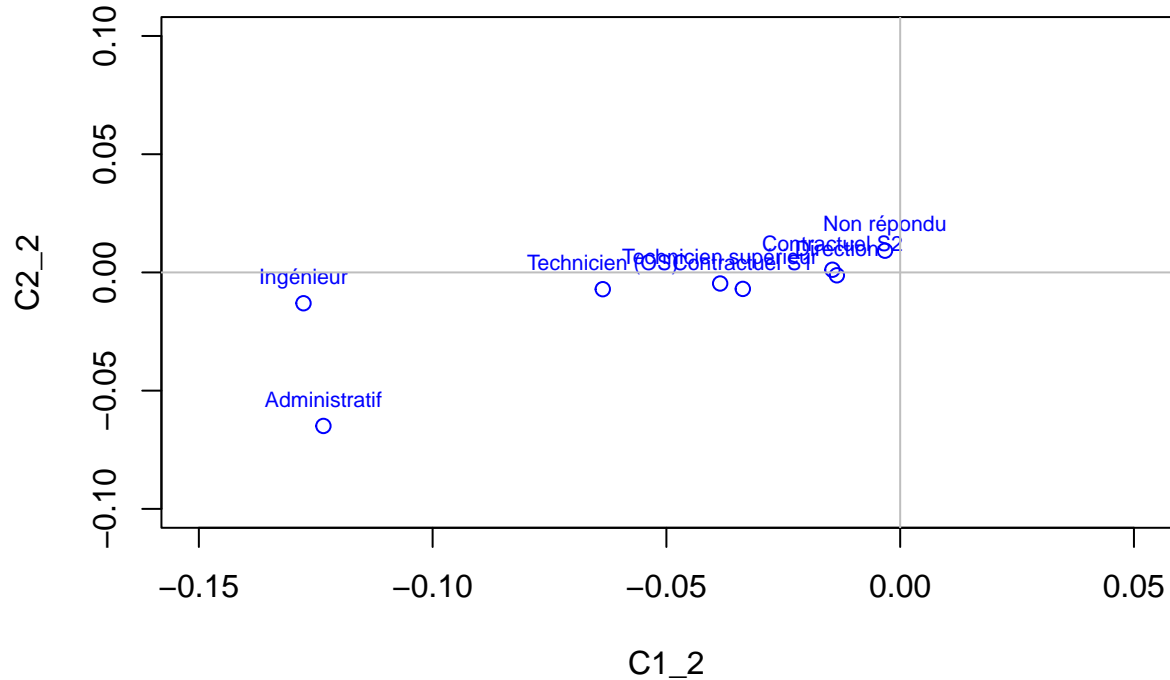
```
# Visualisation du nuage de points dans l'hyperplan obtenu
```

```
plot(C1_2, C2_2, type = "p", xlim = c(-0.15, 0.05), ylim = c(-0.1, 0.1), col = "blue")
```

```
text(C1_2, C2_2, colnames(V), cex = 0.7, pos = 3, col = "blue")
```

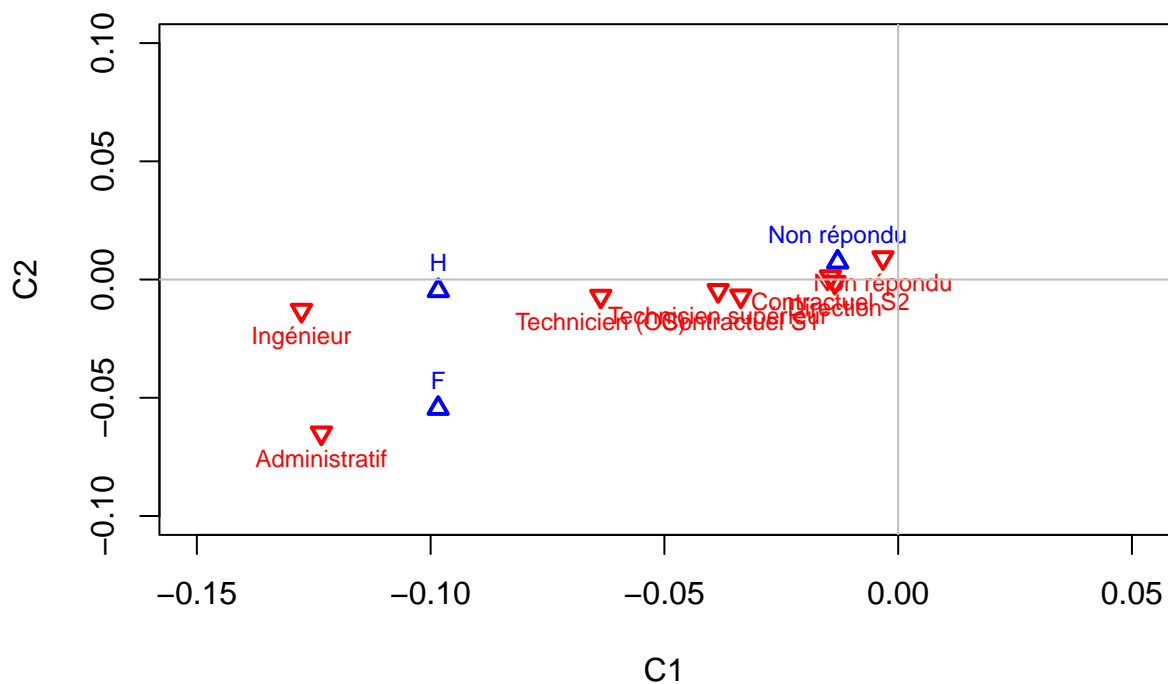
```
abline(h = 0, col = "gray")
```

```
abline(v = 0, col = "gray")
```



3)

```
#All points in the same plot
plot(C1, C2, pch = 24, cex = 1, col = "blue", lwd = 2, xlim = c(-0.15, 0.05), ylim = c(-0.1, 0.1))
text(C1, C2, rownames(V), cex = 0.75, col = "blue", pos = 3)
points(C1_2, C2_2, pch = 25, cex = 1, col = "red", lwd = 2)
text(C1_2, C2_2, colnames(V), cex = 0.75, col = "red", pos = 1)
abline(h = 0, col = "gray")
abline(v = 0, col = "gray")
```



D'après la figure, on voit bien que les hommes ont tendance à occuper la fonction d'ingénieur, alors que les femmes préfèrent plutôt les fonctions administratives. On remarque aussi une très forte ressemblance entre "Direction" et "Contractuel S2", ainsi qu'entre "Technicien supérieur" et "Contractuel S1".

Pour l'interprétation, il est utile de repartir de la répartition du nuage de points et d'en déduire l'indépendance et l'inertie des axes. Une inertie faible entraîne un nuage concentré autour du centre de gravité, alors qu'une inertie grande dilate le nuage. Ici, on remarque que les points sont trop concentrés autour de l'axe de gravité, ce qui implique un taux d'inertie trop faible.

L'indépendance donne un nuage de forme sphérique ; l'existence de dépendance provoque un étirement du nuage dans une direction donnée.

4)

```

# Nouvelles coordonnées dans Rp
co_lignes2 <- V %*% sqrt(Dp) %*% eigen_vectors1

# Nouvelles coordonnées dans Rn
co_colonnes2 <- t(V) %*% sqrt(Dn) %*% eigen_vectors2

quality <- function(n, i, k) {
  A <- sum((n[i, ]^2))
  B <- sum((n[i, 1:k]^2))
  return(B / A)
}

# Qualité des projections sur Rn
quality1 <- rep(0, 3)

for (i in 1:3) {
  quality1[i] <- quality(co_lignes2, i, 2)
}

# Qualité des projections sur Rp
quality2 <- rep(0, 7)

for (i in 1:7) {
  quality2[i] <- quality(co_colonnes2, i, 2)
}

print(quality1)

```

Calcul de la qualité de la projection avec la cascade de valeurs propres, et projection des profils lignes)

```
## [1] 0.7081227 0.6973963 0.7351358
```

```
print(quality2)
```

```
## [1] 0.9060468 0.8335340 0.7940699 0.7967518 0.8895758 0.6277462 0.8498105
```

5)

```
#### Comparaison avec la CA de FactoMineR)
```

```

library("FactoMineR")
library("factoextra")
library("datasets")
library("ggplot2")
library("gplots")
library("dplyr")

```

```
dt <- as.table(as.matrix (V0))
colnames(dt) <- c('0','1','2','3','4','5','6','7')
balloonplot(t (dt), main = "Table de contingence", xlab = "fonction", ylab = "sexe", label = TRUE, show
```

Table de contingence

		fonction							
		0	1	2	3	4	5	6	7
sexe	Non répondu	8		3	7	3			4
	H		15	24	48	13	6	12	4
	F		57	9	18	7	1	6	3
		8	72	36	73	23	7	18	11

Représentation graphique)

```
chisq <- chisq.test(V0)
chisq
```

Test de chi2)

```
##
## Pearson's Chi-squared test
##
## data:  V0
## X-squared = 146.18, df = 14, p-value < 2.2e-16
```

```
res.ca <- CA(V0, graph = FALSE)
```

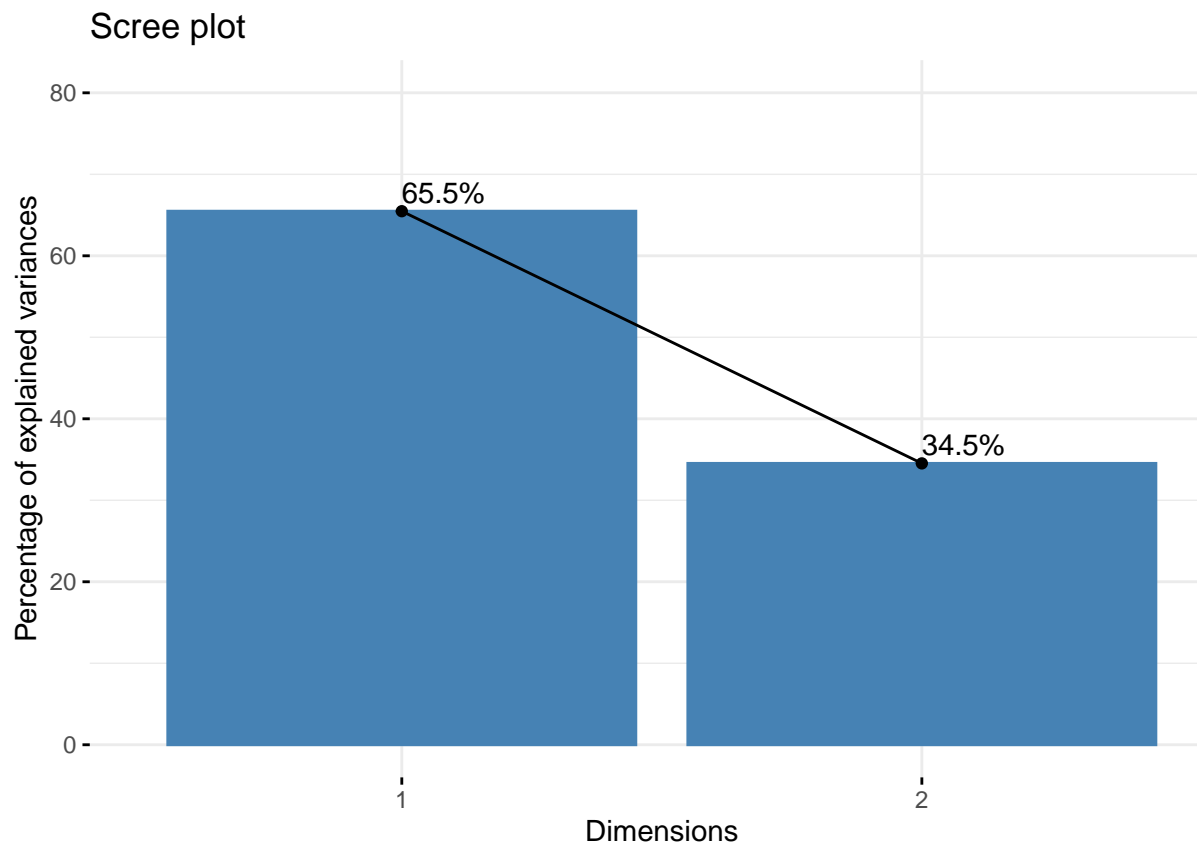
Application de l'AFC)

```
eig.val <- get_eigenvalue(res.ca)
eig.val
```

calcul des inerties(valeurs propres))

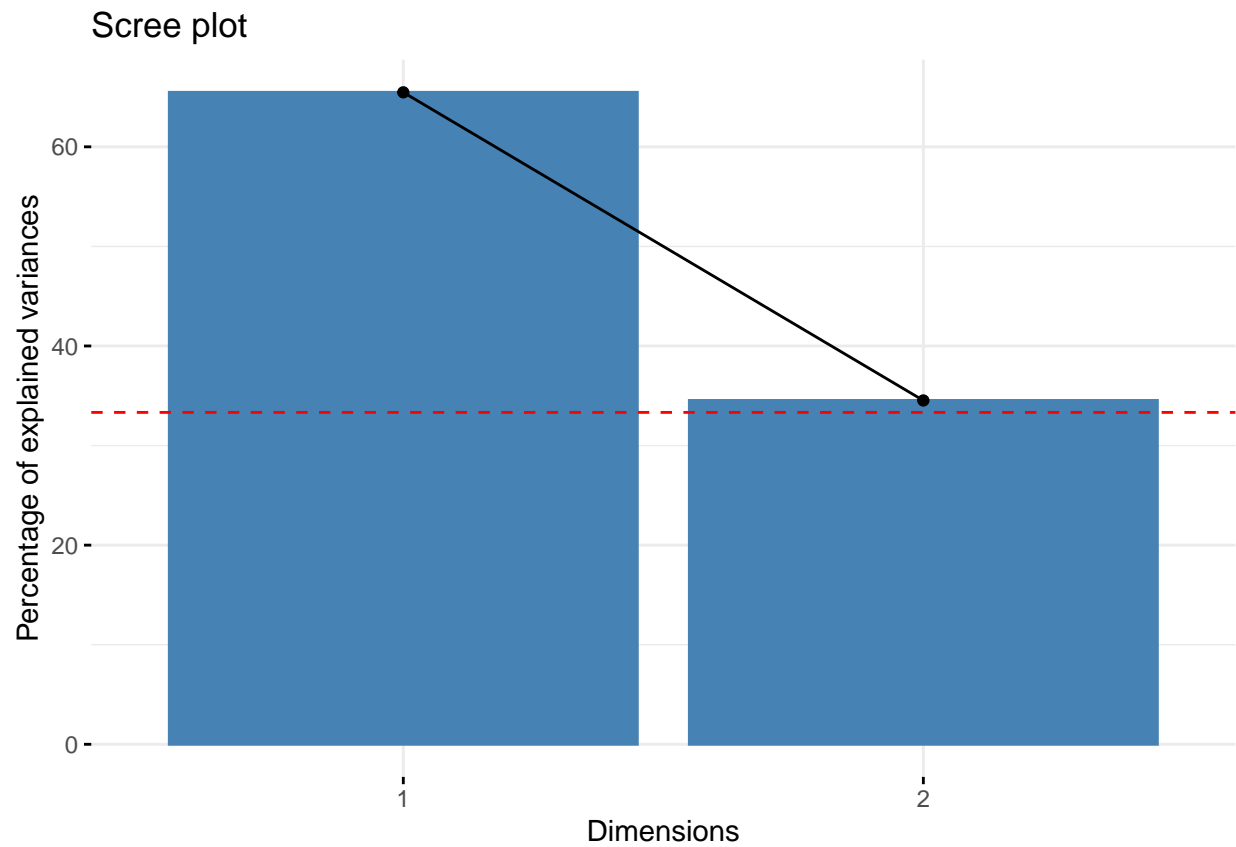
```
##      eigenvalue variance.percent cumulative.variance.percent
## Dim.1  0.3859159         65.47154             65.47154
## Dim.2  0.2035248         34.52846            100.00000
```

```
fviz_screepplot(res.ca, addlabels = TRUE, ylim = c(0, 80))
```



inertie des axes)

```
fviz_screepplot(res.ca) +
  geom_hline(yintercept = 33.33, linetype = 2, color = "red")
```



```
#### Nuage mixte des points)  
fviz_ca_biplot (res.ca, repel = TRUE)
```