Homework 6

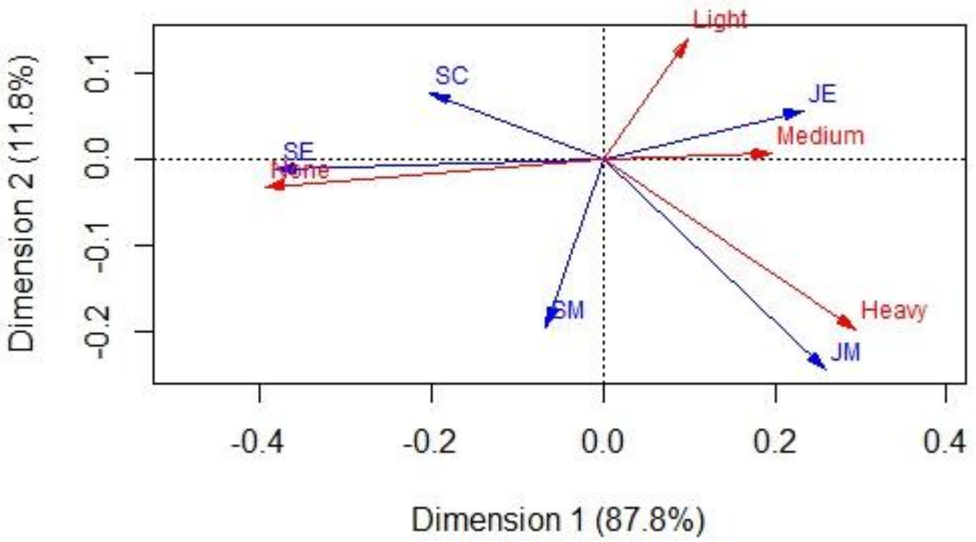
a)

```
> prop.table(S,1) #Row profiles
        None     Light    Medium      Heavy
SM 0.3636364 0.1818182 0.2727273 0.18181818
JM 0.2222222 0.1666667 0.3888889 0.22222222
SE 0.4901961 0.1960784 0.2352941 0.07843137
JE 0.2045455 0.2727273 0.3750000 0.14772727
SC 0.4000000 0.2400000 0.2800000 0.08000000

> prop.table(S,2) #Column profiles
         None      Light    Medium Heavy
SM 0.06557377 0.04444444 0.0483871  0.08
JM 0.06557377 0.06666667 0.1129032  0.16
SE 0.40983607 0.22222222 0.1935484  0.16
JE 0.29508197 0.53333333 0.5322581  0.52
SC 0.16393443 0.13333333 0.1129032  0.08
```

b)

```
> E #Theoretical frequencies under independence
          [,1]      [,2]      [,3]      [,4]
[1,]  3.476684  2.564767  3.533679  1.424870
[2,]  5.689119  4.196891  5.782383  2.331606
[3,] 16.119171 11.891192 16.383420  6.606218
[4,] 27.813472 20.518135 28.269430 11.398964
[5,]  7.901554  5.829016  8.031088  3.238342

> AR.matrix #Attraction Repulsion Matrix
        None     Light    Medium     Heavy
SM 1.1505216 0.7797980 0.8489736 1.4036364
JM 0.7030965 0.7148148 1.2105735 1.7155556
SE 1.5509482 0.8409586 0.7324478 0.6054902
JE 0.6471684 1.1696970 1.1673387 1.1404545
SC 1.2655738 1.0293333 0.8716129 0.6176000

> pchisq(w,df=((I-1)*(J-1)),lower.tail=F)
[1] 0.1718348
# H0: Position and Smoking are independent
> chisq.test(S)

	Pearson's Chi-squared test

data:  S
X-squared = 16.442, df = 12, p-value = 0.1718
```

```
# We can conclude that there is no obvious association betw
een the position and smoking
> S.ca

 Principal inertias (eigenvalues):
           1        2        3
Value      0.074759 0.010017 0.000414
Percentage 87.76%   11.76%   0.49%


 Rows:
              SM        JM        SE       JE        SC
Mass     0.056995  0.093264  0.264249 0.455959  0.129534
ChiDist   0.216559  0.356921  0.380779 0.240025  0.216169
Inertia   0.002673  0.011881  0.038314 0.026269  0.006053
Dim. 1  -0.240539  0.947105 -1.391973 0.851989 -0.735456
Dim. 2  -1.935708 -2.430958 -0.106508 0.576944  0.788435


 Columns:
             None    Light   Medium     Heavy
Mass     0.316062 0.233161 0.321244  0.129534
ChiDist   0.394490 0.173996 0.198127  0.355109
Inertia   0.049186 0.007059 0.012610  0.016335
Dim. 1  -1.438471 0.363746 0.718017  1.074445
Dim. 2  -0.304659 1.409433 0.073528 -1.975960
```



From the above figure, we can get:

Smoking is more frequent among Junior Managers (The angle between “JM” and “Heavy” is very small, the angle between “JM” and “Medium”<90°, the angle between “JM” and “Light” and the angle between “JM” and “None” are >90°).

Smoking is also frequent among Junior Employees (The angle between “JE” and “Medium” is very small, the angle between “JE” and “Heavy”<90°, the angle between “JE” and “None” are >90°).

Smoking is less frequent among Senior Employees and Secretaries (The angle between “SE” and “None” is very small, the angle between “SE” and “Heavy”, the angle between “SE” and “Medium”, and the angle between “SE” and “Light” are >90°).

For Senior Managers, there is no obvious conclusion. There may be an extreme smoking type distribution: Some don’t like smoking at all, while some others are heavy smoker (The angle between ”SM” and “None” and the angle between “SM” and “Heavy” <90°, while the angle between “SM” and “Light”, and the angle between “SM” and “Medium”>90°).

c)

To sum up, we can get: the results are in harmony with home exercise 5.

Appendix

Code:

```
setwd("C:/Users/tracy/Desktop/Multivariate Statistical Analysis/作业/作业 6/directory")
install.packages("ca")
library(ca)
data<-read.table("SMOKING.txt",header=T,row.names=1)
dim(data)
View(data)
S<-as.matrix(data[-6,-5])
prop.table(S,1) #Row profiles
prop.table(S,2) #Column profiles
prop.table(S) #Table of relative frequencies
v1 <- matrix(colSums(S),nrow=1)
v2 <- matrix(rowSums(S),ncol=1)
n<-sum(S)
E <- v2 %*% v1/n #Theoretical frequencies under independence
AR.matrix<-S/E #Attraction Repulsion Matrix

I <- dim(S)[1]
J <- dim(S)[2]
w <- 0
for(i in 1:I){
  for(j in 1:J){
    w <- w + ( S[i,j]-E[i,j] )^2/(E[i,j])
  }
```

```
}
pchisq(w,df=((I-1)*(J-1)),lower.tail=F)
chisq.test(S)
S.ca <- ca(S)
names(S.ca)
S.ca$sv
S.ca$rowdist
S.ca$rownames
S.ca$coldist
S.ca$colnames
S.ca$rowcoord
S.ca$colcoord
S.ca$rowinertia
S.ca$colinertia
sum(S.ca$rowinertia)
sum(S.ca$colinertia)
sum(S.ca$sv^2)
w/n
S.ca$rowinertia/sum(S.ca$sv^2)
S.ca$rowmass
S.ca$colmass
margin.table(as.matrix(S),1)/sum(S)
margin.table(as.matrix(S),2)/sum(S)
S.ca$N
S.ca
summary(S.ca)
names(summary(S.ca)$rows)
plot(S.ca)
plot(S.ca,arrows=c(T,T),map="symmetric")
plot(S.ca,arrows=c(T,T),map="symmetric",dim=c(1,2))
install.packages("rgl")
plot3d.ca(S.ca)
```