

## Homework 5

a)

Code:

```
setwd("C:/Users/tracy/Desktop/Multivariate Statistical Analysis/作业/作业 5/directory")
smoking <- read.table("SMOKING.txt",header=T,row.names = 1)
View(smoking)
S <- as.matrix(smoking[,-5])
Total<-margin.table(S)
prop.table(S) #Table of relative frequencies
prop.table(S,1) #Row profiles
prop.table(S,2) #Column profiles
RowNum <- matrix(margin.table(S,1),ncol=1)
RowPro<-prop.table(RowNum)
ColNum<- matrix(margin.table(S,2),nrow=1)
ColPro<-prop.table(ColNum)
TPUI<-RowPro%*%ColPro #Form theoretical proportions under independence
TFUI<-Total*TPUI      #Form theoretical frequencies under independence
matrix(margin.table(TFUI,1),ncol=1)
matrix(margin.table(TFUI,2),nrow=1)
# The following also can obtain the result.
#V1 <- matrix(margin.table(S,1),ncol=1)
#V2 <- matrix(margin.table(S,2),nrow=1)
#E <- V1 %*% V2/sum(S)
AR.matrix<-S/TFUI
```

	[,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

b) 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

c)

From the attraction repulsion matrix, we can get:

Smoking is more frequent among Junior Managers (The values of "Heavy" and "Medium"  $>1$  and they are the biggest, and the values of "None" and "Light"  $<1$ ).

Smoking is also frequent among Junior Employees (The values of "Heavy" , "Medium" and "Light" are all a little bigger than 1 , and the values of "None"  $<1$ ).

Smoking is less frequent among Senior Employees and Secretaries (The value of "None"  $>1$ , and almost all the values of other smoking categories are  $<1$ ).

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 value of "None" and "Heavy"  $>1$ , while the values of "Light" and "Medium"  $<1$ ).