HW5. for Multivariate Statistics II

November 17, 2020

201611531/Department of Statistics/Jeong Hojae

Chapter 8. Correspondence Analysis(CRA) : Multiple CRA

The following data are the three-way table for the car accident record on the highway depending on whether the safety belt is worn.

[Table 1]

안전벨트	이탈	사망	
		예	아니오
착용	예	14	1
	아니오	483	411
미착용	예	497	462
	아니오	100	157

(1) Explain category variables and classification variables in [Table 1].

categorical variables : 안전벨트, 이탈, 사망

classification variables : 착용/미착용, 예/아니오, 예/아니오

(2) Make a data form of classification variables from [Table 1].

```
> accident
  이탈 사망 안전벨트 Freq
1  yes  yes     착용   14
2   no  yes     착용  483
3  yes   no     착용    1
4   no   no     착용  411
5  yes  yes   미착용  497
6   no  yes   미착용  100
7  yes   no   미착용  462
8   no   no   미착용  157
```

(3) Apply multiple CRA to the data in (2).

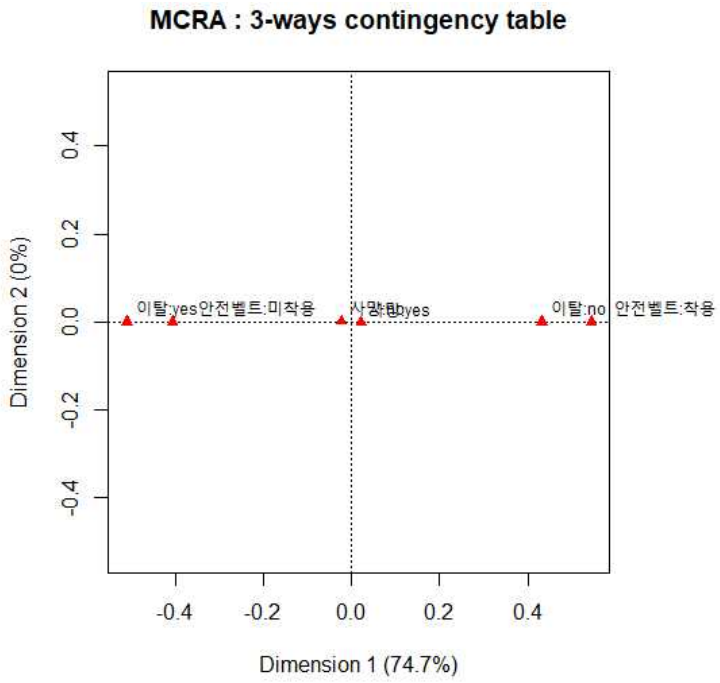
```
> mjca(table)
```

```
 Eigenvalues:
            1          2
Value       0.147308 2e-06
Percentage  74.75%    0%
```

```
 Columns:
            이탈:no    이탈:yes    사망:no    사망:yes    안전벨트:미착용    안전벨트:착용
Mass        0.180549   0.152784   0.161725   0.171608         0.190745        0.142588
ChiDist     0.669315   0.790946   0.595732   0.561426         0.629616        0.842259
Inertia     0.080883   0.095581   0.057396   0.054091         0.075615        0.101152
Dim. 1      1.124677  -1.329058  -0.061194   0.057670        -1.059518        1.417352
Dim. 2      0.114342  -0.135121   1.779101  -1.676648         0.034830       -0.046593
```

Dim1's GOF is 74.75%. Dim2's GOF is 0%. It can be shown the associations between classification variables by Dim1.

(4) Describe some associations between categorical variables in the multiple CRA map.



Relative positions of the points : similarities and differences among rows and columns categories.

"이탈:yes" and "안전벨트:미착용", "사망:no" are related to each other because they have same direction.
"이탈:no" and "안전벨트:착용", "사망:yes" are related to each other because they have same direction.

We can see that wearing the seat belt is related to lane departure and death.

2. [Exercise 8.10] (p.501)

(1) Apply multiple CRA

Burt matrix

```
> B
      교육1 교육2 교육3 교육4 적성1 적성2 적성3 적성4 적성5 직업1 직업2 직업3 직업4
 [1,]  1040     0     0     0   215   281   372   128    44   239     6     1   794
 [2,]     0  1108     0     0   208   285   386   176    53   309    11     7   781
 [3,]     0     0  1237     0   138   284   446   238   131   233    70    12   922
 [4,]     0     0     0   968    83   197   385   186   117    53   199   215   501
 [5,]   215   208   138    83   644     0     0     0     0   122    30    20   472
 [6,]   281   285   284   197     0  1047     0     0     0   226    51    66   704
 [7,]   372   386   446   385     0     0  1589     0     0   306   115    96  1072
 [8,]   128   176   238   186     0     0     0   728     0   130    59    38   501
 [9,]    44    53   131   117     0     0     0     0   345    50    31    15   249
[10,]   239   309   233    53   122   226   306   130    50   834     0     0     0
[11,]     6    11    70   199    30    51   115    59    31     0   286     0     0
[12,]     1     7    12   215    20    66    96    38    15     0     0   235     0
[13,]   794   781   922   501   472   704  1072   501   249     0     0     0  2998
```

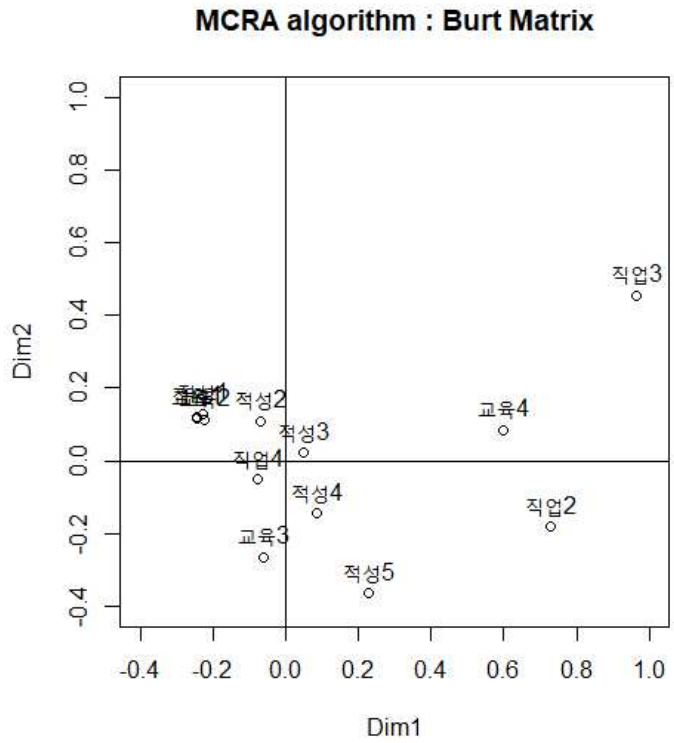
Coordinate

```
> Cb2
              [,1]        [,2]
교육1 -0.24275450  0.11799191
교육2 -0.22574360  0.11321747
교육3 -0.06306682 -0.26602887
교육4  0.59979570  0.08359625
적성1 -0.22720684  0.13062126
적성2 -0.06906785  0.10974704
적성3  0.04903160  0.02106607
적성4  0.08463271 -0.14601324
적성5  0.22930848 -0.36580175
직업1 -0.24599999  0.12024326
직업2  0.72887452 -0.18158516
직업3  0.96369167  0.45393605
직업4 -0.07663831 -0.05170930
```

GOF

```
> rbind(round(lam, 3),round(fit, 3))
        [,1]   [,2]   [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]  [,10] [,11] [,12] [,13]
[1,]   0.273  0.148  0.126 0.116 0.112 0.109 0.104 0.093 0.082 0.024     0     0     0
[2,] 23.039 12.500 10.620 9.779 9.425 9.160 8.799 7.810 6.885 1.983     0     0     0
```

(2) What is GOF of multiple CRA map. And interpret that.



Dim1's GOF is 23.039%. Dim2's GOF is 12.5%.

Relative positions of the points : similarities and differences among rows and columns categories.

"적성3~5" and "교육4", "직업2~3" are related to each other because they have same direction.

We can see that high aptitude levels is related to high educational levels. In this situation, these are related to professional and educational professions.

3. Consider the [Data 8.5.1] (view.txt) Economic Outlook and Economic Policy Preference Survey Data.

(1) Apply multiple CRA to the data

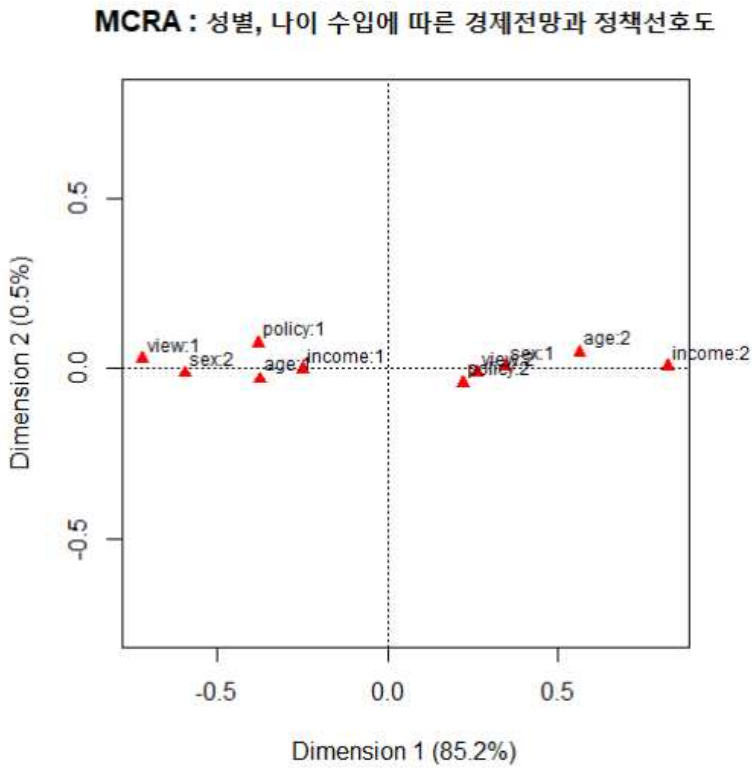
```
> head(view)
   sex age income view policy
1    1   2      2    2      2
2    1   1      1    1      2
3    2   1      1    1      1
4    1   2      2    2      2
5    2   1      1    2      2
6    1   2      1    2      1
> mjca(view)

 Eigenvalues:
                1          2
Value      0.178447 0.001018
Percentage 85.18%   0.49%


 Columns:
             sex:1      sex:2      age:1      age:2   income:1 income:2
Mass     0.126667   0.073333   0.120000 0.080000   0.153333 0.046667
ChiDist 0.465567   0.804162   0.528152 0.792229   0.341826 1.123143
Inertia 0.027455   0.047423   0.033473 0.050210   0.017916 0.058868
Dim. 1   0.815374 -1.408374 -0.888077 1.332115 -0.591738 1.944283
Dim. 2   0.230783 -0.398625 -0.984092 1.476138 -0.068827 0.226146
             view:1     view:2     policy:1   policy:2
Mass     0.053333   0.146667   0.073333   0.126667
ChiDist  0.998656   0.363147   0.718208   0.415804
Inertia  0.053190   0.019342   0.037827   0.021900
Dim. 1  -1.701435   0.618704 -0.896624   0.519098
Dim. 2   0.906130 -0.329502   2.329320 -1.348553
```

Dim1's GOF is 85.18%. Dim2's GOF is 0.49%. It can be shown the associations between classification variables by Dim1.

(2) Describe some associations between categorical variables in the multiple CRA map.



Relative positions of the points : similarities and differences among rows and columns categories.

"age:1" and "income:1", "sex:2", "policy:1", "income:1" are related to each other because they have same direction.
"age:2" and "income:2", "sex:1", "policy:2", "income:2" are related to each other because they have same direction.

We can see that women who have under the age of 30 and under the income of 10 million won are optimistic about the economic outlook and the government's economic policies.
In contrast, men who have over the age of 30 and over the income of 10 million won are pessimism about the economic outlook and the government's economic policies.

```
setwd("E:/학교/2020 2학기 정호재/다변량통계학2/실습/20201117/Rdata")
Freq <-array(c(14,483,1,411,497,100,462,157))
이탈<-c("yes", "no")
사망<- c("yes", "no")
안전벨트<-c("착용","미착용")
data<-expand.grid(이탈=이탈, 사망=사망, 안전벨트=안전벨트)
accident<-cbind(data, Freq)
accident

table<-xtabs(Freq~이탈+사망+안전벨트, data=accident)
table

#install.packages("ca")
library(ca)
mjca(table)
plot(mjca(table),main="MCRA : 3-ways contingency table")
##############################################
apitude_burt<-read.table("apitude.txt",header=T)

B<-as.matrix(apitude_burt)
B
P <- B / sum(B)
round(P,3)
cm <- apply(P, 2, sum)
Dc<-diag(1/sqrt(cm))
eP <- cm %*% t(cm)
Y <- (P - eP) / sqrt(eP)
round(Y,3)

# Singular Value Decomposition
svd.Y<-svd(Y)
V<-svd.Y$v
Dl<-diag((svd.Y$d)^2)

lam<-(svd.Y$d)^2
fit<-lam/sum(lam)*100
rbind(round(lam, 3),round(fit, 3))

Cb<- Dc%*%V%*%Dl
rownames(Cb)<-colnames(apitude_burt)

Cb2<-Cb[, 1:2]
```

```
Cb2


lim<-range(pretty(Cb))

plot(Cb2, xlab="Dim1", ylab="Dim2", xlim=lim, ylim=lim, main="MCRA algorithm : Burt Matrix")
text(Cb2, colnames(apitude_burt), pos=3, col=1)
abline(v=0, h=0)

##########################################
view<-read.table("view.txt",header=T)
view<-view[,-1]
head(view)

library(ca)
mjca(view)
par(pty="s")
plot(mjca(view), main="MCRA : 성별, 나이 수입에 따른 경제전망과 정책선호도")
```