2016판

한남대학교 통계학과 권세혁교수

Multiple CA(대응분석)

속성이 3개 이상인 경우 - 개체+(범주형변수) MCA()함수

```r
# load packages
library(FactoMineR)
library(ggplot2)

# load data tea
data(tea)

# select these columns
newtea = tea[, c("Tea", "How", "how", "sugar", "where", "always")]
head(newtea)
```

```
> head(newtea)
        Tea   How      how    sugar      where    always
1     black alone  tea bag    sugar chain store Not.always
2     black  milk  tea bag No.sugar chain store Not.always
3 Earl Grey alone  tea bag No.sugar chain store Not.always
4 Earl Grey alone  tea bag    sugar chain store Not.always
5 Earl Grey alone  tea bag No.sugar chain store     always
6 Earl Grey alone  tea bag No.sugar chain store Not.always
```

```r
# number of categories per variable
cats = apply(newtea, 2, function(x) nlevels(as.factor(x)))
cats
```

```r
# apply MCA
mca1 = MCA(newtea, graph = FALSE)
mca1
mca1$eig #table of eigenvalues

# data frame with variable coordinates
mca1_vars_df = data.frame(mca1$var$coord, Variable = rep(names(cats),
cats))

# data frame with observation coordinates
mca1_obs_df = data.frame(mca1$ind$coord)

# plot of variable categories
ggplot(data=mca1_vars_df,
    aes(x = Dim.1, y = Dim.2, label = rownames(mca1_vars_df))) +
geom_hline(yintercept = 0, colour = "gray70") +
geom_vline(xintercept = 0, colour = "gray70") +
geom_text(aes(colour=Variable)) +
  ggtitle("MCA plot of variables using R package FactoMineR")
```

MCA plot of variables using R package FactoMineR

MCA plot of variables using R package FactoMineR
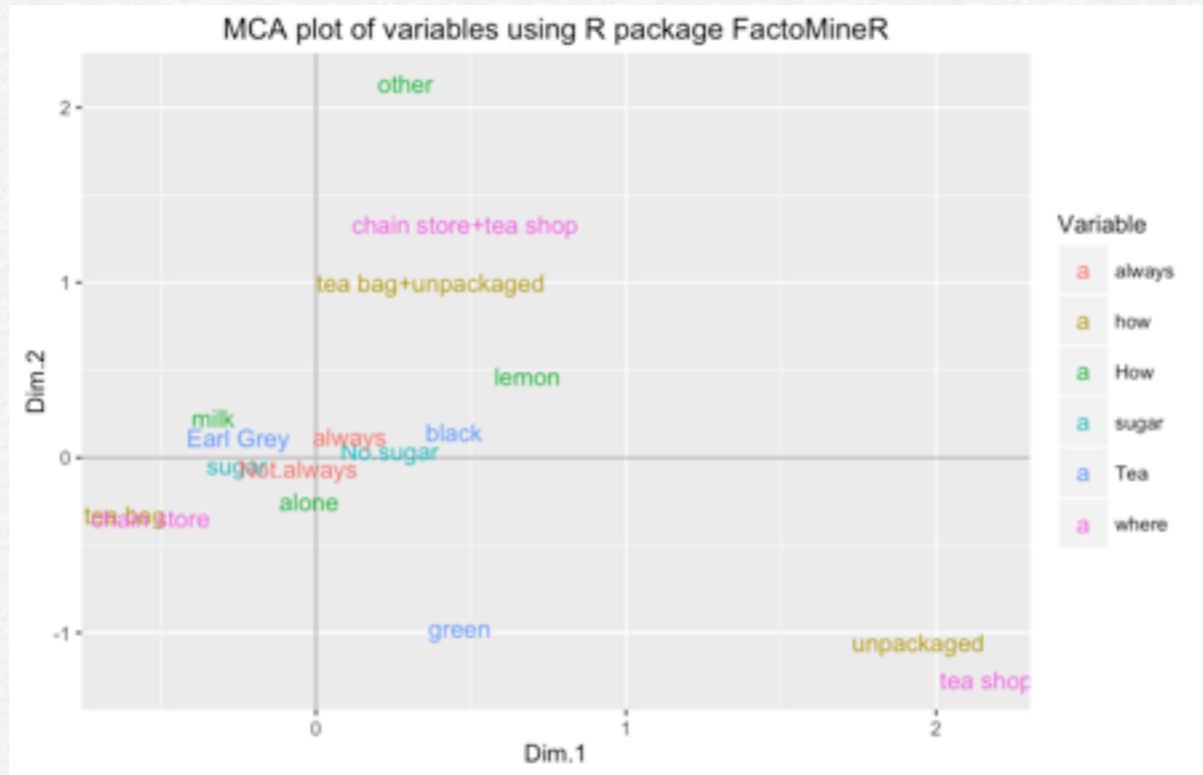
MCA plot of observations and categories

```
ggplot(data = mca1_obs_df, aes(x = Dim.1, y = Dim.2)) +
  geom_hline(yintercept = 0, colour = "gray70") +
  geom_vline(xintercept = 0, colour = "gray70") +
  geom_point(colour = "gray50", alpha = 0.7) +
  geom_density2d(colour = "gray80") +
  geom_text(data = mca1_vars_df,
        aes(x = Dim.1, y = Dim.2,
          label = rownames(mca1_vars_df), colour = Variable)) +
  ggtitle("MCA plot of variables using R package FactoMineR") +
```
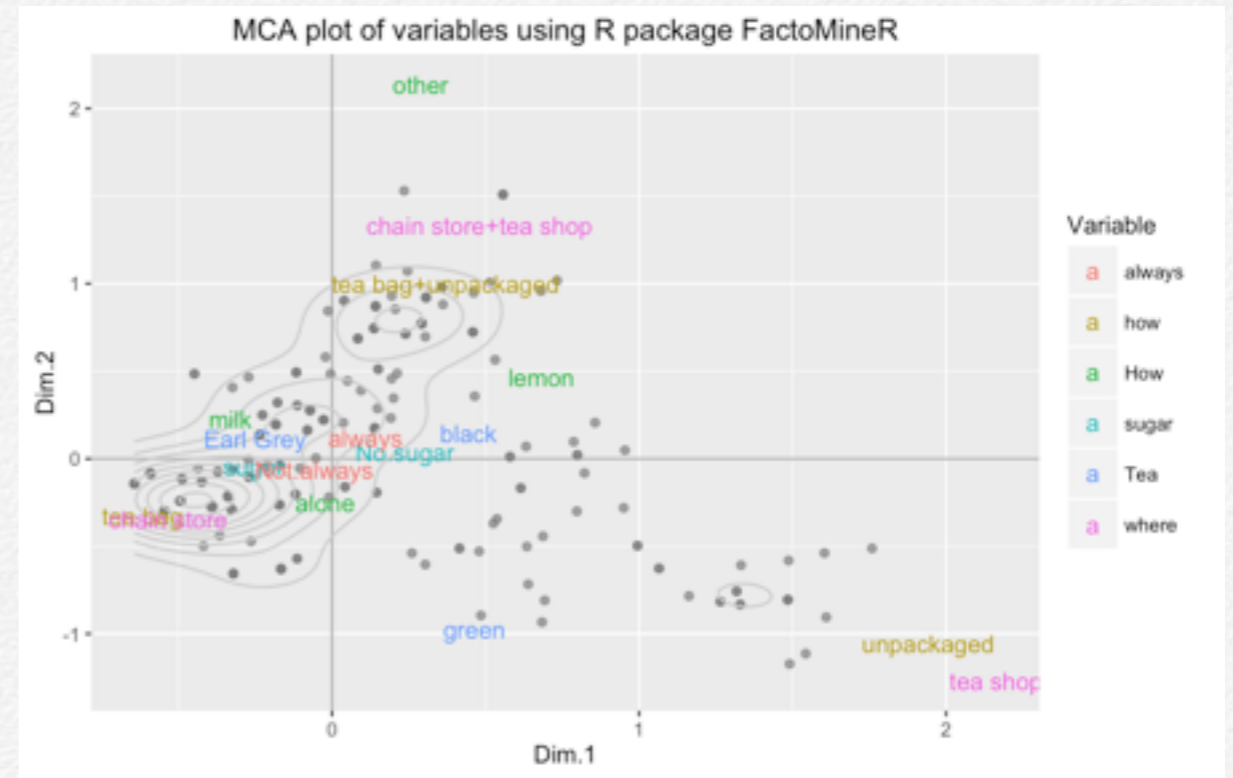
*) 개체 – 성별*연령*소득수준

*) 범주형 변수 – 유병여부, 아침식사여부, 비만, **여가생활 유형**

속성이 3개 이상인 경우 - 개체+(범주형변수) —mca()함수

```
# load MASS
library(MASS)
# apply mca
mca2 = mca(newtea, nf = 5)
# eigenvalues
```

mca2$d^2

[1] 0.2798 0.2577 0.2201 0.1879 0.1688

column coordinates

head(mca2$cs)

```
> mca2$d^2
[1] 0.2797618 0.2577477 0.2201379 0.1879296 0.1687650
> head(mca2$cs)
                         1            2            3            4            5
Tea.black      -0.0081110752  0.002718746 -0.023117893  0.016378641  0.004040303
Tea.Earl Grey   0.0045537734  0.002112519  0.009847745 -0.001549967 -0.001838069
Tea.green      -0.0084442029 -0.018451615 -0.005754263 -0.027662904  0.001689843
How.alone       0.0003981841 -0.004760162 -0.002117646 -0.010152078 -0.005461704
How.lemon      -0.0124131888  0.008792981  0.025994990  0.022239227 -0.027419175
How.milk        0.0060214907  0.004332055 -0.001424218  0.014219535  0.031529737
```

row coordiantes

head(mca2$rs)

```
> head(mca2$rs)
           1            2            3            4            5
1 0.003145138 -0.002792722 -0.0036675526  0.003906394 -0.002255623
2 0.002590013 -0.001018727 -0.0076950149  0.005356172  0.004107402
3 0.003763603 -0.002635135 -0.0023163134 -0.001693865 -0.003037567
4 0.005255946 -0.002893760  0.0018267203  0.000918293 -0.003235352
5 0.003258968 -0.002068720  0.0003119625 -0.003609338  0.002300557
6 0.003763603 -0.002635135 -0.0023163134 -0.001693865 -0.003037567
```

data frame for ggplot

mca2_vars_df = data.frame(mca2$cs, Variable = rep(names(cats), cats))

plot

ggplot(data = mca2_vars_df,

    aes(x = X1, y = X2, label = rownames(mca2_vars_df))) +

geom_hline(yintercept = 0, colour = "gray70") +

geom_vline(xintercept = 0, colour = "gray70") +

geom_text(aes(colour = Variable)) +

ggtitle("MCA plot of variables using R package MASS")