

**2018.5.21**

## 1 문제 1

dt8 에서 하나의 양적 변수를 선택해서 2 개의 수준을 갖는 질적 변수로 변환시키시오.  
이변수의 이름을 res 라고 하시오. res 수준이 높으면 1 낮으면 0 이라고 놓으시오.

```
load(file="C:/2018-2/wo/homework324/dt8.RData")
#v18-자산총액
dt8$res=ifelse(dt8$V18<mean(dt8$V18),0,1)
table(dt8$res)
##
##      0      1
## 12650  5623
```

## 2 문제 2

res 가 1 이 될 확률을 추정해 줄 만한 질적 설명변수 1 개를 골라서 이 변수 하나를  
가지고 res 가 1 이 될 확률을 추정하시오.

```
#설명변수-v103 가구주 교육정도
table(dt8$V103)
##
## A1201 A1202 A1203 A1204
##   3425   1971   6080   6797
```

```
dt8$rgroup <- runif(n = nrow(dt8))
```

```
dTrainAll <- subset(dt8,rgroup<=0.9)
dTest <- subset(dt8,rgroup>0.9)
```

```
useForCal <-
  rbinom(n=dim(dTrainAll)[[1]],size=1,prob=0.1) > 0
head(useForCal,20)
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE
```

```
dCal <- subset(dTrainAll,useForCal)
dTrain <- subset(dTrainAll,!useForCal)
table(dTrain$V103)
##
## A1201 A1202 A1203 A1204
##   2782   1603   4974   5439
```

```
dT1 = dTrain[dTrain$V103 == 'A1201', ]
td1 = table(dT1$res)
```

```
a1 = td1[2]/sum(td1)
a1
##          1
## 0.1024443
```

```
dT2 = dTrain[dTrain$V103=='A1202', ]
td2 = table(dT2$res)
a2 = td2[2]/sum(td2)
a2
##          1
## 0.1996257
```

```
dT3 = dTrain[dTrain$V103 == 'A1203', ]
td3 = table(dT3$res)
a3 = td3[2]/sum(td3)
a3
##          1
## 0.2808605
```

```
dT4 = dTrain[dTrain$V103== 'A1204',]
td4 = table(dT4$res)
a4 = td4[2]/sum(td4)
a4
##          1
## 0.4653429
```

```
tdT = table(dTrain$res)
tdT
##
##      0      1
## 10265  4533
```

```
pt = tdT[2]/sum(tdT)
pt
##          1
## 0.3063252
```

```
dTrain$pred = pt
dTrain[dTrain$V103 == 'A1201' , ]$pred = a1
dTrain[dTrain$V103 == 'A1202' , ]$pred = a2
dTrain[dTrain$V103 == 'A1203' , ]$pred = a3
dTrain[dTrain$V103 == 'A1204' , ]$pred = a4
table(dTrain$pred)
##
##      0.102444284687275      0.199625701809108      0.28086047446723
## 0.465342893914322
##           2782           1603           4974           543
9
```