

건강실태조사

조서연

2020 9 29

1.데이터 정리

```
data <- read.csv("Rawdata.csv")

health <- data[,c(3,4,12,18,39,44,45,89,159,161,245,259,260,275)]
names(health) <- c("age","sex","legion_code","family","height","weight","smoke","alchol","stress","depress",
"physical","inter1","inter2","job")

health$sex <- factor(health$sex,levels=c(1,2),labels=c("남자","여자"))

health$smoke[health$smoke==1] <- 0
health$smoke[health$smoke==2] <- 0
health$smoke[health$smoke==3] <- 1
health$smoke[health$smoke==7] <- 1
health$smoke[health$smoke==9] <- 1
health$smoke <- factor(health$smoke,levels=c(0,1),labels=c("핀다","안핀다"))

health$alchol[health$alchol==7] <- 2
health$alchol[health$alchol==9] <- 2
health$alchol <- factor(health$alchol,levels=c(1,2),labels=c("마신다","안마신다"))

health$stress[health$stress==7] <- NA
health$stress[health$stress==9] <- NA
health$stress <- factor(health$stress,levels=c(1,2,3,4),labels=c("대단히 많이 느낀다","많이 느끼는 편이다","조금
느끼는 편이다","거의 느끼지 않는다"))

health$depress[health$depress==7] <- NA
health$depress[health$depress==9] <- NA
health$depress <- factor(health$depress,levels=c(1,2),labels=c("예","아니오"))

health$physical[health$physical==7] <- NA
health$physical[health$physical==9] <- NA

health$inter1[health$inter1==7] <- NA
health$inter1[health$inter1==9] <- NA
health$inter1 <- factor(health$inter1,levels=c(1,2,3,4,5,6),labels=c("한달에1번미만","한달에1번","한달에2~3번","
일주일에1번","일주일에2~3번","일주일에4번이상"))
health$inter2[health$inter2==7] <- NA
health$inter2[health$inter2==9] <- NA
health$inter2 <- factor(health$inter2,levels=c(1,2,3,4,5,6),labels=c("한달에1번미만","한달에1번","한달에2~3번","
일주일에1번","일주일에2~3번","일주일에4번이상"))

health$job[health$job==7] <- NA
health$job[health$job==9] <- NA
health$job <- factor(health$job,levels=c(1,2),labels=c("예","아니오"))

health_data <- health[health$age>=65,]
Data <- health_data[health_data$physical<=2,]

write.csv(Data,file="localhealth.csv",row.names = TRUE)
```

2.우울증에 미치는 요인 분석(1)

```
library(gmodels)
data <- read.csv("localhealth.csv")
data <- na.omit(data)
data$BMI <- data$weight/(data$height/100)^2
summary(data)
```

```
##      X              age              sex              legion_code
## Length:72977      Min.    : 65.00      Length:72977      Min.    :11110
## Class :character  1st Qu.: 69.00      Class :character  1st Qu.:41117
## Mode  :character  Median : 74.00      Mode  :character  Median :44150
##                               Mean  : 74.46
##                               3rd Qu.: 79.00
##                               Max.   :105.00
##                               Max.   :50130
##      family      height      weight      smoke
## Min.    :10.00      Min.    : 50.6      Min.    : 21.9      Length:72977
## 1st Qu.:10.00      1st Qu.: 150.0      1st Qu.: 52.0      Class :character
## Median :11.00      Median : 156.8      Median : 59.1      Mode  :character
## Mean    :14.34      Mean    : 4060.3      Mean    : 2465.8
## 3rd Qu.:11.00      3rd Qu.: 164.8      3rd Qu.: 66.7
## Max.    :99.00      Max.    :99999.0      Max.    :99999.0
##      alcohol      stress      depress      physical
## Length:72977      Length:72977      Length:72977      Min.    :1.0
## Class :character  Class :character  Class :character  1st Qu.:1.0
## Mode  :character  Mode  :character  Mode  :character  Median :1.0
##                               Mean    :1.4
##                               3rd Qu.:2.0
##                               Max.    :2.0
##      inter1      inter2      job      BMI
## Length:72977      Length:72977      Length:72977      Min.    : 0.00
## Class :character  Class :character  Class :character  1st Qu.: 21.50
## Mode  :character  Mode  :character  Mode  :character  Median : 23.75
##                               Mean    : 92.91
##                               3rd Qu.: 26.00
##                               Max.    :286505.36
```

#로지스틱 회귀 분석

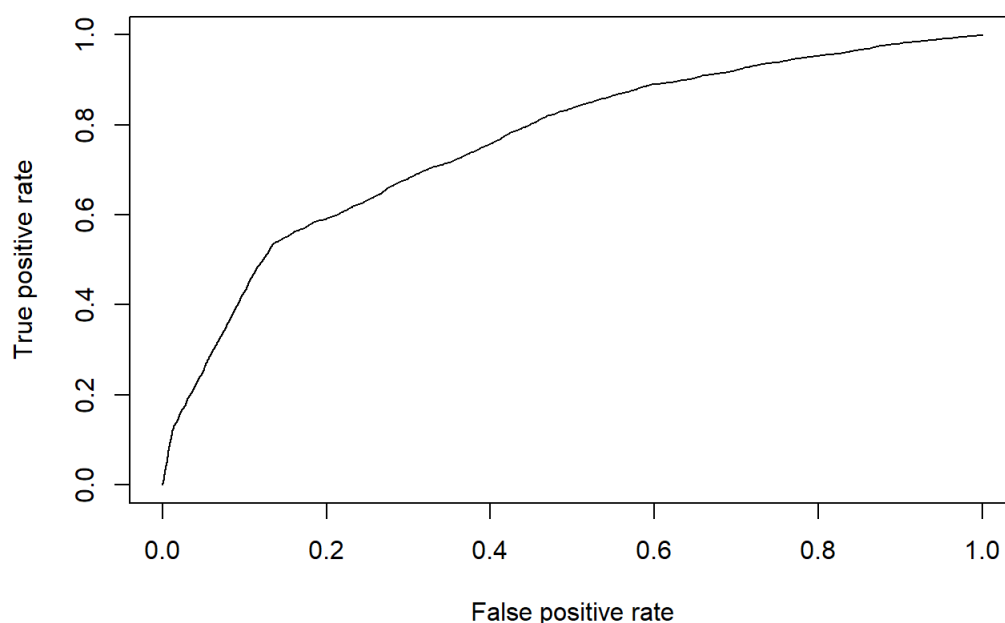
```
data$depress[data$depress=="아니오"] <- 0
data$depress[data$depress=="예"] <- 1
data$stress <- factor(data$stress,levels=c("대단히 많이 느낀다","많이 느끼는 편이다","조금 느끼는 편이다","거의 느끼지 않는다"),labels=c(1,2,3,4))
data$inter1 <- factor(data$inter1,levels=c("한달에1번미만","한달에1번","한달에2~3번","일주일에1번","일주일에2~3번","일주일에4번이상"),labels=c(1,2,3,4,5,6))
data$inter2 <- factor(data$inter2,levels=c("한달에1번미만","한달에1번","한달에2~3번","일주일에1번","일주일에2~3번","일주일에4번이상"),labels=c(1,2,3,4,5,6))
data$stress <- as.factor(data$stress)
data$inter1 <- as.factor(data$inter1)
data$inter2 <- as.factor(data$inter2)
data$depress <- as.numeric(data$depress)

model <- glm(depress ~ smoke + alcohol + stress + BMI, data = data, family = "binomial")
summary(model)
```

```
##
## Call:
## glm(formula = depress ~ smoke + alchol + stress + BMI, family = "binomial",
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0575  -0.3419  -0.3059  -0.2081   3.0959
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.131e-01  5.587e-02  -5.603 2.10e-08 ***
## smoke핀다    -2.285e-01  3.590e-02  -6.366 1.95e-10 ***
## alchol안마신다 2.438e-02  3.409e-02   0.715  0.4745
## stress2      -1.036e+00  5.810e-02 -17.840 < 2e-16 ***
## stress3      -2.497e+00  5.813e-02 -42.953 < 2e-16 ***
## stress4      -3.508e+00  6.725e-02 -52.159 < 2e-16 ***
## BMI          -2.131e-05  1.290e-05  -1.652  0.0986 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 36095  on 72976  degrees of freedom
## Residual deviance: 31252  on 72970  degrees of freedom
## AIC: 31266
##
## Number of Fisher Scoring iterations: 6
```

#0.05보다 작은 변수로는 흡연유무, 스트레스 정도가 영향을 미치는 변수이다.
 # 그중 스트레스는 estimate값이 음수이므로 스트레스 랭크가 낮아질수록(스트레스를 더 많이 받을 수록) 우울증에 걸릴 확률이 높다.
 #bmi는 연관이 없다.

```
library(ROCR)
p <- predict(model, newdata=data, type="response")
pr <- prediction(p, data$depress)
prf <- performance(pr, measure = "tpr", x.measure = "fpr")
plot(prf)
```



```
auc <- performance(pr, measure = "auc")
auc <- auc@y.values[[1]]
auc
```

```
## [1] 0.7610812
```

#모델의 auc값이 0.76으로 모델로서 적합하다.

3.공원면적이 우울증에 미치는 영향

```
data <- read.csv("localhealth.csv")
data <- na.omit(data)
data$legion_code <- as.character(data$legion_code)
summary(data$legion_code)
```

```
##      Length      Class      Mode 
##      72977 character character
```

```
data$region[data$legion_code<20000] <- "서울시"
data$region[data$legion_code>20000&data$legion_code<27000] <- "부산시"
data$region[data$legion_code>27000&data$legion_code<28000] <- "대구시"
data$region[data$legion_code>28000&data$legion_code<29000] <- "인천시"
data$region[data$legion_code>29000&data$legion_code<30000] <- "광주시"
data$region[data$legion_code>30000&data$legion_code<31000] <- "대전시"
data$region[data$legion_code>31000&data$legion_code<32000] <- "울산시"
data$region[data$legion_code>32000&data$legion_code<33000] <- "세종시"
data$region[data$legion_code>33000&data$legion_code<34000] <- "경기도"
data$region[data$legion_code>34000&data$legion_code<35000] <- "강원도"
data$region[data$legion_code>35000&data$legion_code<36000] <- "충청북도"
data$region[data$legion_code>36000&data$legion_code<37000] <- "충청남도"
data$region[data$legion_code>37000&data$legion_code<38000] <- "전라북도"
data$region[data$legion_code>38000&data$legion_code<39000] <- "전라남도"
data$region[data$legion_code>39000&data$legion_code<40000] <- "경상북도"
data$region[data$legion_code>40000&data$legion_code<41000] <- "경상남도"
data$region[data$legion_code>41000] <- "제주특별자치도"

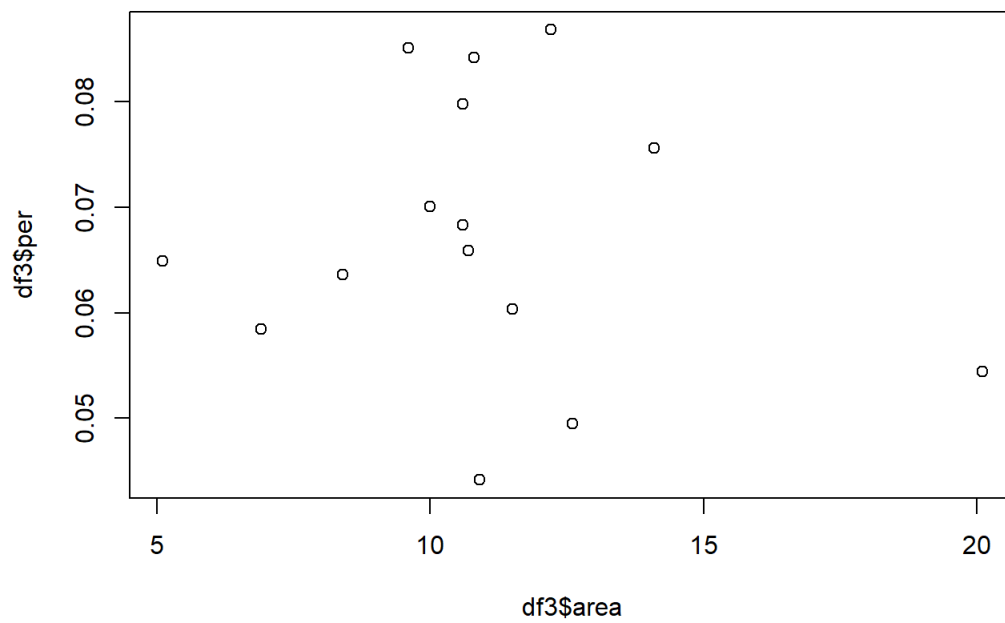
summary.factor(data$region)
```

##	강원도	경기도	경상남도	경상북도	광주시	대구시	대전시	
##	5766	8810	6462	8710	1163	1833	1087	
##	부산시	서울시	세종시	울산시	인천시	전라남도	전라북도	
##	4330	5718	176	863	2411	9007	5675	
##	제주특별자치도	충청남도	충청북도					
##	1606	5149	4211					

```
per <- c(
  (length(which(data$region=="서울시" & data$depress=="예"))/length(which(data$region=="서울시"))),
  (length(which(data$region=="부산시" & data$depress=="예"))/length(which(data$region=="부산시"))),
  (length(which(data$region=="대구시" & data$depress=="예"))/length(which(data$region=="대구시"))),
  (length(which(data$region=="울산시" & data$depress=="예"))/length(which(data$region=="울산시"))),
  (length(which(data$region=="대전시" & data$depress=="예"))/length(which(data$region=="대전시"))),

  (length(which(data$region=="경기도" & data$depress=="예"))/length(which(data$region=="경기도"))),
  (length(which(data$region=="강원도" & data$depress=="예"))/length(which(data$region=="강원도"))),
  (length(which(data$region=="충청북도" & data$depress=="예"))/length(which(data$region=="충청북도"))),
  (length(which(data$region=="충청남도" & data$depress=="예"))/length(which(data$region=="충청남도"))),
  (length(which(data$region=="경상북도" & data$depress=="예"))/length(which(data$region=="경상북도"))),
  (length(which(data$region=="광주시" & data$depress=="예"))/length(which(data$region=="광주시"))),
  (length(which(data$region=="경상남도" & data$depress=="예"))/length(which(data$region=="경상남도"))),
  (length(which(data$region=="전라남도" & data$depress=="예"))/length(which(data$region=="전라남도"))),
  (length(which(data$region=="전라북도" & data$depress=="예"))/length(which(data$region=="전라북도"))),
  (length(which(data$region=="제주" & data$depress=="예"))/length(which(data$region=="제주"))),
  (length(which(data$region=="인천시" & data$depress=="예"))/length(which(data$region=="인천시")))
)

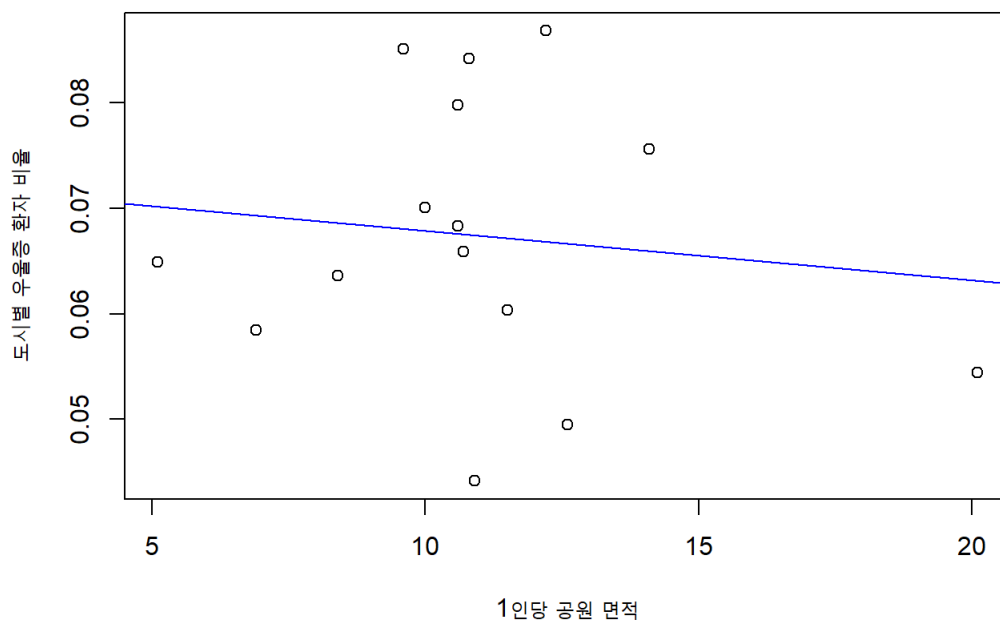
area <- c(8.4,12.2,5.1,10.6,10.9,9.6,10.7,10.6,10.0,11.5,6.9,12.6,20.1,14.1,6.0,10.8)
df3 <- data.frame(per,area)
plot(x=df3$area,y=df3$per)
```



```
m1 <- lm(df3$per~df3$area)
m1
```

```
##
## Call:
## lm(formula = df3$per ~ df3$area)
##
## Coefficients:
## (Intercept)      df3$area
##    0.072549    -0.000468
```

```
plot(df3$per~df3$area,xlab="1인당 공원 면적",ylab="도시별 우울증 환자 비율")
abline(m1,col='blue')
```



```
summary(m1)
```

```
##
## Call:
## lm(formula = df3$per ~ df3$area)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.023290 -0.007759 -0.001639  0.010923  0.019996
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.072550   0.012188   5.953  4.8e-05 ***
## df3$area     -0.000468   0.001068  -0.438   0.668
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01344 on 13 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.01456,    Adjusted R-squared:  -0.06125
## F-statistic: 0.192 on 1 and 13 DF,  p-value: 0.6684
```

#유의확률이 0.05이상으로 설명 불가->모델로서 적합성 없음