[8주차][2017011794][김연수]통계학 실습

1. 코드

DF<-read.csv('week8.csv')

DF

tapply(DF$Effect,DF$Medicine,mean)

tapply(DF$Problem,DF$Medicine,mean)

DF2<-tapply(DF$Problem,DF$Medicine,mean)

sort(DF2,decreasing = T)

DF3<data.frame(Medicine=c("A","B","C"),Effect\_prob=c(0.566,0.274,0.390),Problem\_prob=c(0.454,0.396,0.286))

DF3

DF4<-c(0.112,-0.396,0.104)

DF5<-cbind(DF3,DF4)

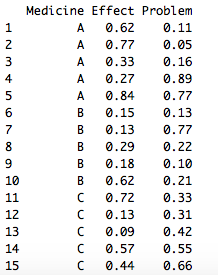
DF5

1. 코드 해석   
   - 데이터 프레임 처리

DF<-read.csv('week8.csv')

DF

Csv파일을 데이터 프레임으로 읽기 위해 read.csv() 함수를 이용하였다.



-약물 별 평균 효과 발생률 구하기

tapply(DF$Effect,DF$Medicine,mean)

데이터 내 Factor의 범주에 의해 그룹화 되어 해당 그룹별로 함수를 적용해야

한다는 판단 하에 tapply함수를 이용하였다.



-약물 별 평균 문제 발생률 구하기

tapply(DF$Problem,DF$Medicine,mean)

이 역시 위와 같은 이유로 tapply함수를 이용하였다.



-문제가 발생할 확률이 높은 약물부터 출력

DF2<-tapply(DF$Problem,DF$Medicine,mean)

sort(DF2,decreasing = T)

먼저, 문제가 발생될 확률이 높은 약물부터 순서대로 출력해야함으로 먼저 문제 발생률 판단한 값을 DF2에 저장해 두었다. 그런 다음 sort함수를 이용해서 내림차순으로 그 값을 출력하였다.



-어떤 약물을 사용해야 하는지 판별하기

DF3<data.frame(Medicine=c("A","B","C"),Effect\_prob=c(0.566,0.274,0.390),Problem\_prob=c(0.454,0.396,0.286))

DF3

DF4<-c(0.112,-0.396,0.104)

DF5<-cbind(DF3,DF4)

DF5

어떤 약물이 효과적인지 판별하기 위해서는 약물이 실제적으로 효과적인지 판별해야한다. 그렇게 하기 위해서는 (효과 발생률-문제 발생률)의 값을 비교해야 실제적 약의 효능을 알 수 있다. 효과가 아무리 높아도 문제 발생률이 높으면, 부작용과 같은 문제가 발생하기 때문에 고려해주어야 한다.

먼저 위에서 구한 효과 발생률, 문제 발생률 데이터들을 dataframe화 시켜준다. 그리고 나서 (효과 발생률-문제 발생률) 값을 포함한 DF4 데이터를 DF3에 붙여주었다. 그 결과 A의 실제적 효과 발생률이 0.112로 가장 높다는 것을 알게 되었고, 사업을 위해서는 A의 약물을 사용해야 함을 알 수 있었다.

