

Linear Regression & Logistic Regression 복습 문제

김서희 | 김세정 | 김장미 | 우현우

다음은 R 내장 데이터 state.x77를 데이터프레임으로 만든 states입니다.
states는 살인사건 발생률, 인구수, 문맹률 등을 담고 있습니다.

```
states <- as.data.frame(state.x77)
colnames(states)
```

1-1.

states 데이터를 활용하여 **살인사건 발생률(Murder)**을 종속변수로 하는 다중 선형 회귀 모델 'fit'을 만들고, '단계적 선택법'을 활용하여 최종적으로 선택된 변수들을 확인하고, 새로운 모델 'fit2'를 만들어 주세요.

1-2.

'fit2'에서 최종적으로 선택된 설명변수들은 무엇인가요?

1-3.

인구(population)과 면적(area)의 베타 계수를 해석해 보세요.

'사례연구.csv' 파일을 이용하도록 하자.

해당 파일은..

독립변수로서 '과목subject', '나이age'을 가지고 종속변수로서 '선택한 수업방식teaching'을 가진다.

(이때, teaching값이 discuss일 때는, “토론식 수업을 선택한다”, lecture일 때에는 “강의형 수업을 선택한다” 라고 해석하자)

이를 로지스틱 회귀모델로 만들고, 다음에 대해 답하여라.

2-1. 각 변수별로 **오즈비(Odds Ratio) 수치** 값을 구하라. (단, 모든 변수는 편의상 Factor형으로 하자)

(원래라면, age는 ordina이지만, 본 과제에서는 X, Y가 모두 범주형에 한하여 로지스틱회귀를 진행하기 위함임)

2-2. 1에서 구한 오즈비(Odds Ratio)를 해석하라. 이때, 해석할 변수는 science과 age11과 age13만 하면 됩니다. (3개만)

***더 생각해볼 사항:** 종속변수이자 y값인 teaching은 discuss와 lecture라는 값이 있는데..

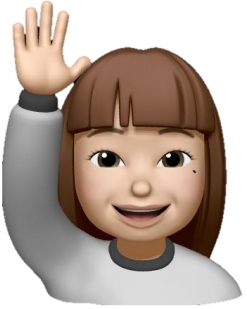
(y=0일때 discuss이고, y=1일때 lecture일까?) 아니면 (y=0일때 lecture이고, y=1일때 discuss일까?)

2-3. '2-2'에서 로지스틱회귀의 오즈비를 구한 결과,

그렇다면 **"intercept(절편값)"**도 다른변수 오즈비(Odds Ratio) **해석에 포함해야할까?** (T/F) (근거도 함께 제출)

(이 문제 역시, X, Y가 모두 범주형categorical에 한정하여 내는 문제라는 점 참고)

복습 문제 풀이에 사용될 데이터 불러오기 !



```
# 준비단계_라이브러리 및 데이터 불러오기
library(foreign)
library(nnet)
library(ggplot2)
library(reshape2)

ml <- read.dta("https://stats.idre.ucla.edu/stat/data/hsbdemo.dta")
with(ml, table(ses, prog))
with(ml, do.call(rbind, tapply(write, prog, function(x) c(M = mean(x), SD = sd(x))))))
```

The data set contains variables on 200 students.

The *response variable* (Y) is prog, program type.

The *predictor variables* (X) are social economic status, ses,
a three-level categorical variable and writing score, write, a continuous variable.

3-1.

데이터의 prog 변수에서 reference 변수를 “academic”으로 재설정된 값을 prog2라는 이름으로 데이터에 변수를 추가하세요 !
그 후, ses와 write 변수를 X변수로 하여 prog2를 다중 로지스틱 회귀로 적합하고, p-value까지 구해보세요 !
(hint : releval() 함수, multinom() 함수)

3-2.

3-1번에서 적합한 회귀계수에 대한 해석을 부탁드립니다 !

(reference 변수를 반영하여 해주시면 됩니다 ! 세션 ppt를 참고하시면 쉬울 것입니다^.^)

Linear Regression &
Logistic Regression
복습 문제 끝 !

김서희 | 김세정 | 김장미 | 우현우