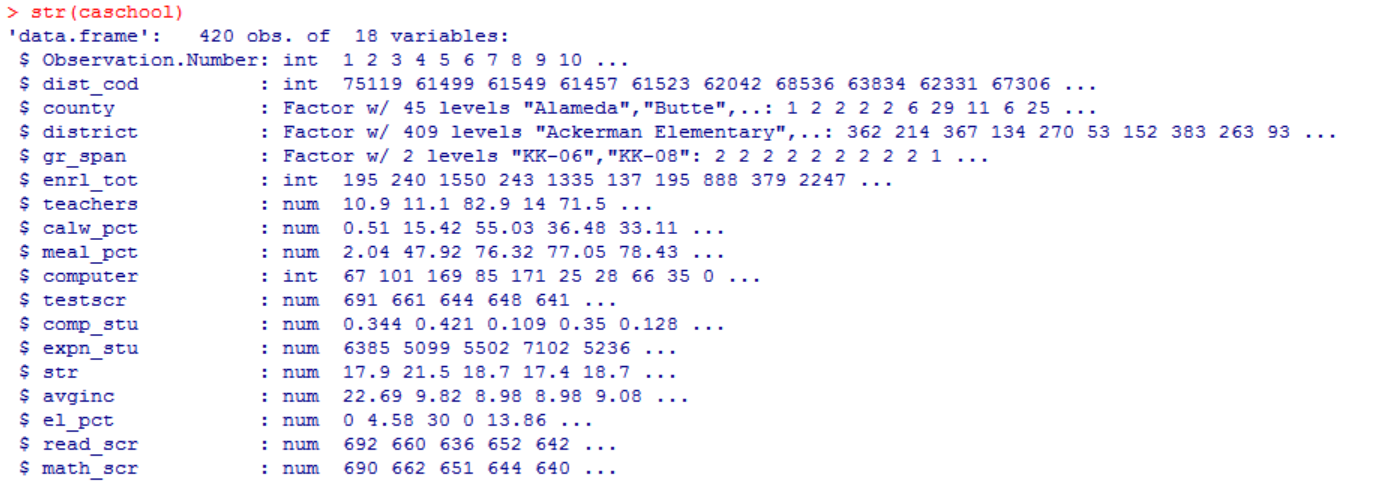
B

1)



Caschool 데이터는 미국 캘리포니아 초등학교 성적관련 자료로, 420개의 관측치, 18개의 변수를 포함한다.이 중 우리가 분석에서 고려하고자 하는 변수 read\_scr, avginc, el\_pct, str, meal\_pct 모두 num으로 연속형 변수이다.

2) 분석의 목적은 교사 1인당 학생수 str을 감소 시킬 때, 이에 따른 교육효과 read\_scr가 어떠한지 파악하기 위함. Str이 외에 성적에 영향을 미치는 다른 경제, 인종 변수들 (Average district income(logarithm), %English learner, %Eligible for subsidized lunch) 고려(통제)해 str의 효과를 알아낸다.

3) 종속변수 read\_scr~



1. lm.fit0 <- lm(read\_scr~str)

STR효과 -2.6210

1. lm.fit1 <- lm(read\_scr~str+meal\_pct+el\_pct)

영어보충생비율, 무료급식비율 통제 후 STR효과 -1.18594

1. lm.fit2 <- lm(read\_scr~str+meal\_pct+el\_pct+log(avginc))

추가적으로 로그소득을 통제 후 STR 효과 -0.92636

1. lm.fit3 <- lm(read\_scr~str+HiEL+str\*HiEL)

STR\*HiEL상호작용의 pvalue=0.1983 으로 유의하지 않다.

***<질문1>*** STR감소로 인한 RS 효과가 %English learner에 영향을 받지 않는다.

1. lm.fit4 <- lm(read\_scr~str+HiEL+str\*HiEL+meal\_pct+log(avginc))

통제변수(무료급식,로그소득), STR\*HiEL 상호작용의 pvalue=0.45561 유의X

1. lm.fit5 <- lm(read\_scr~str+str2+str3+HiEL+meal\_pct+log(avginc))

STR을 변화시키는 효과가 STR의 값에 영향을 받는지 알아보기 위해 STR 3차항까지 포함시키고 모형 4에서 HiEL과 STR의 교호작용이 유의하지 않아 삭제한 모형.

H0:b2=b3=0 의 p-value<0.001보다 작아 기각, |t|=3.390(pr<0.01) , |t|=3.469(pr<0.01)

***<질문2>*** STR이 비선형 효과를 보인다.

1. lm.fit6 <-lm(read\_scr~str+str2+str3+HiEL+HiEL\*str+HiEL\*str2+HiEL\*str3+meal\_pct+log(avginc))

STR, STR^2,STR^3,HiEL,HiEL\*STR, HiEL\*STR^2, HiEL\*STR^3,무료급식률, 로그소득 고려,

HiEL\*STR, HiEL\*STR^2, HiEL\*STR^3: RS와 STR 사이의 회귀함수가 영어를 배우는 학생의 비율이 낮고 높음에 따라 상이한지를 확인. H0: b5=b6=b7=0, P<0.05

***<질문3>*** STR 교사당 2명 줄일 때(경제 요인과 비선형성 고려 후) 이것의 RS에 대한 추정효과

(2) lm.fit2 linear로 기울기가 일정해 STR값에 영향을 받지 않음

(5)(7) 비선형 setting의 경우 STR값에 영향을 받는다.

경제적 요소를 고려한 후 %English learner의 높고 낮음이 STR 변화에 따른 RS의 효과에 실질적인 영향을 주지 못하며, RS에 대한 STR의 비선형 관계가 존재한다. 또한 비선형 setting의 경우 STR값에 따라 RS에 대한 추정효과가 달라진다.

5) R코드

caschool = read.csv("~/caschool.csv",header=TRUE)

attach(caschool)

head(caschool)

str(caschool)

plot(str,read\_scr)

lm.fit0 <- lm(read\_scr~str)

summary(lm.fit0)

abline(lm.fit0,col="red")

lm.fit1 <- lm(read\_scr~str+meal\_pct+el\_pct)

summary(lm.fit1)

lm.fit2 <- lm(read\_scr~str+meal\_pct+el\_pct+log(avginc))

summary(lm.fit2)

HiEL <- ifelse(el\_pct>10, 1, 0)

lm.fit3 <- lm(read\_scr~str+HiEL+str\*HiEL)

summary(lm.fit3)

lm.fit4 <- lm(read\_scr~str+HiEL+str\*HiEL+meal\_pct+log(avginc))

summary(lm.fit4)

str2<-(caschool$str)^2

str3<-(caschool$str)^3

lm.fit5 <- lm(read\_scr~str+str2+str3+HiEL+meal\_pct+log(avginc))

summary(lm.fit5)

lm.fit6 = lm(read\_scr~str+str2+str3+HiEL+HiEL\*str+HiEL\*str2+HiEL\*str3+meal\_pct+log(avginc))

summary(lm.fit6)

lm.fit7 = lm(read\_scr~str+str2+str3+el\_pct+meal\_pct+log(avginc))

summary(lm.fit7)