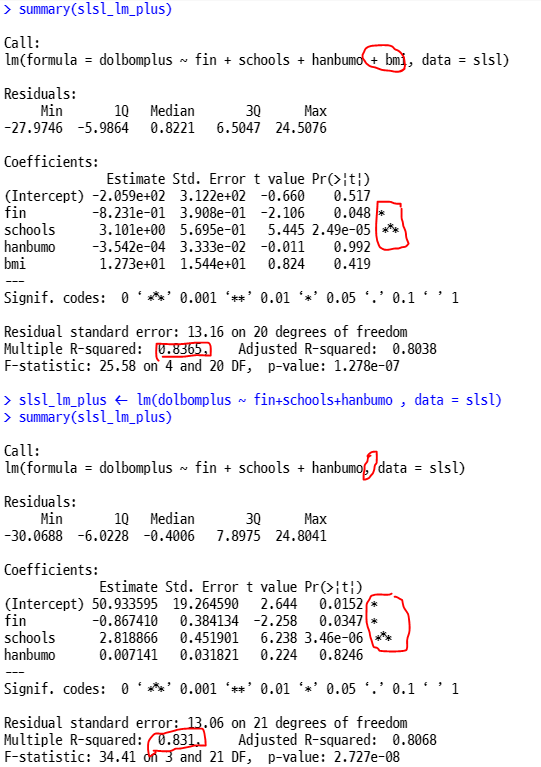
\*지역별 데이터는 가급적 최근연도(2018-2020)로 반영해서 모델에 넣었음

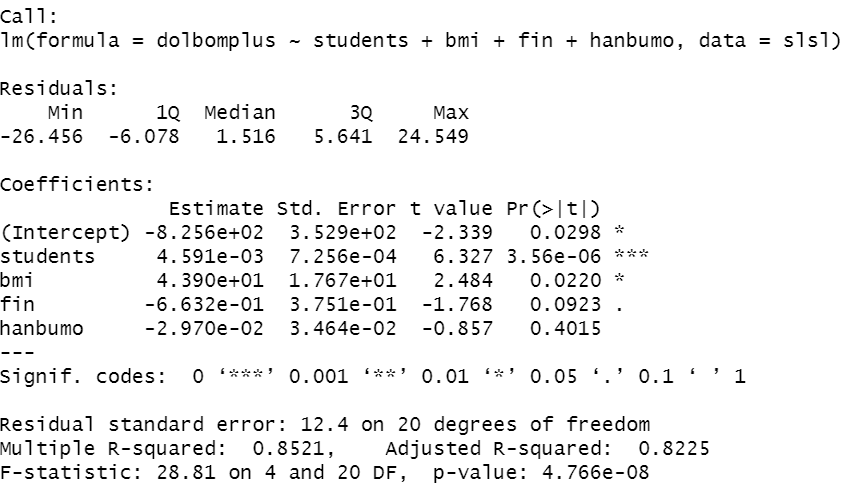
[서울 모델]

-시행착오:돌봄교실 변수 자체가 계산할 때 학교수를 이용하여 계산했기 때문에 당연히 학교수가 관련이 압도적으로 높게 나오는 거라고, 학교수 쓰면 안된다고 결론 내림

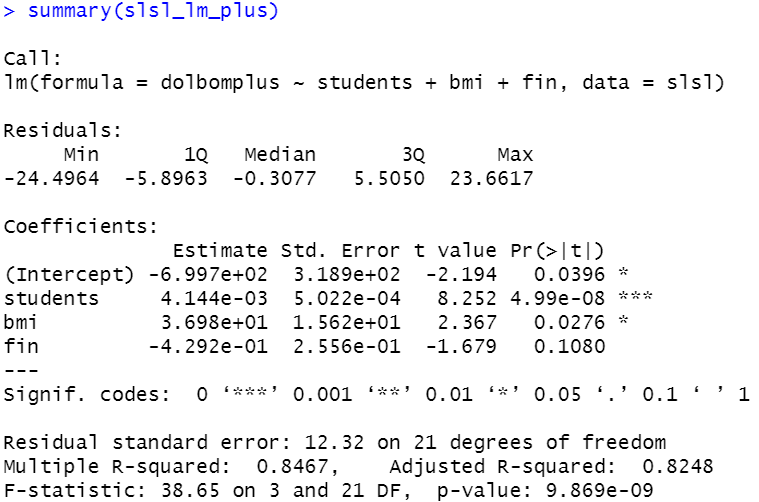


-한부모 넣었을 때 → 재정자립도 의미있게 나옴

돌봄교실 전체 수랑 재정자립도만 따로 봤을 때는 약하지만 음의관계

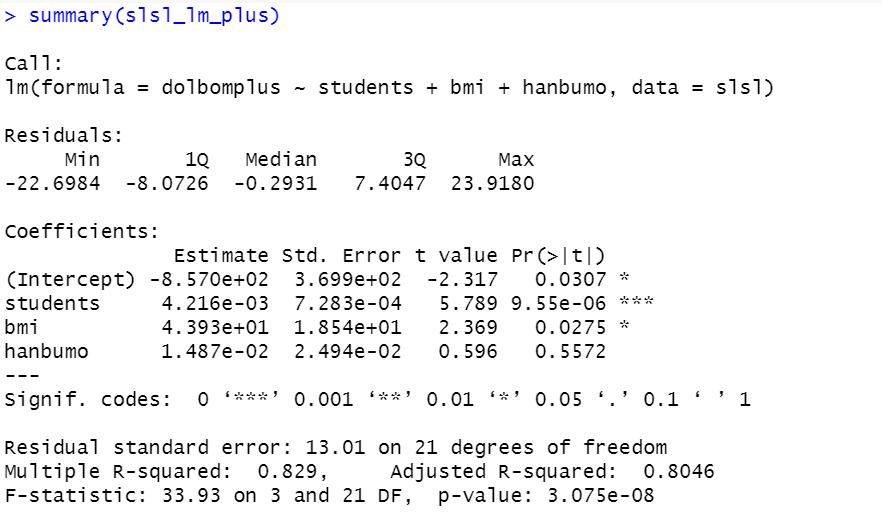


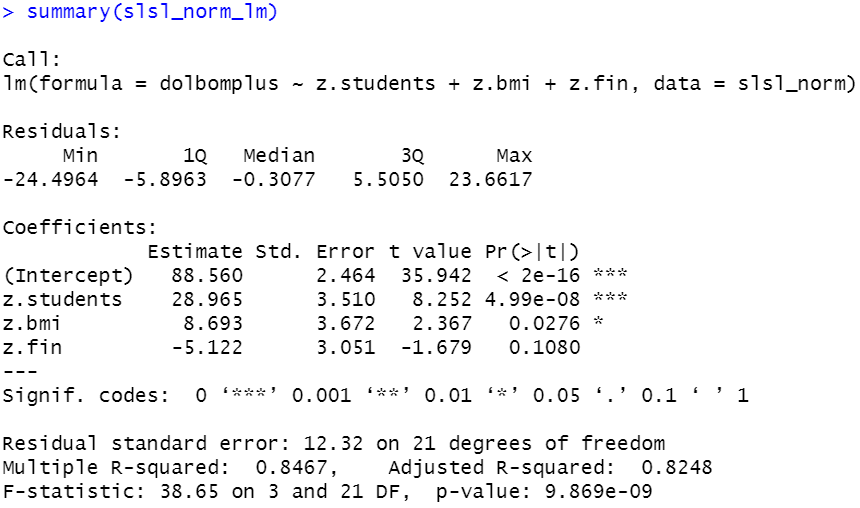
-한부모 제외 시 → 재정자립도 별 없음



-한부모 넣고 재정자립도 뺐을 때 결과: 한부모는 재정자립도 유무에 따라 계수의 부호가 흔들리는 모습

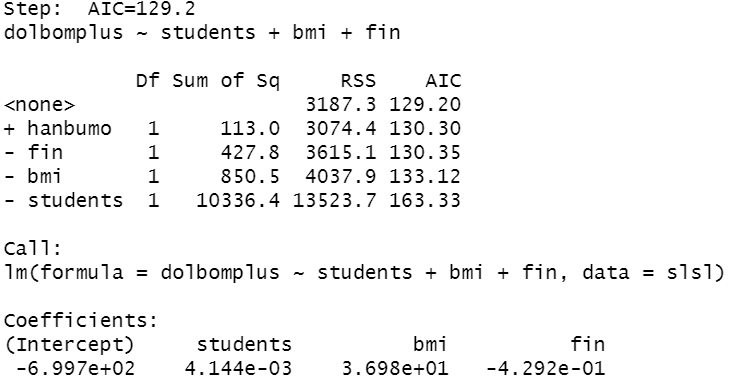
-summary에서는 재정자립도를 빼라고 이야기하긴 하지만, stepwise 기법에서는 재정자립도를 넣는 게 좋다고 하여 한부모 없이 재정자립도를 갖고 가는 방법 취해도 좋을 듯

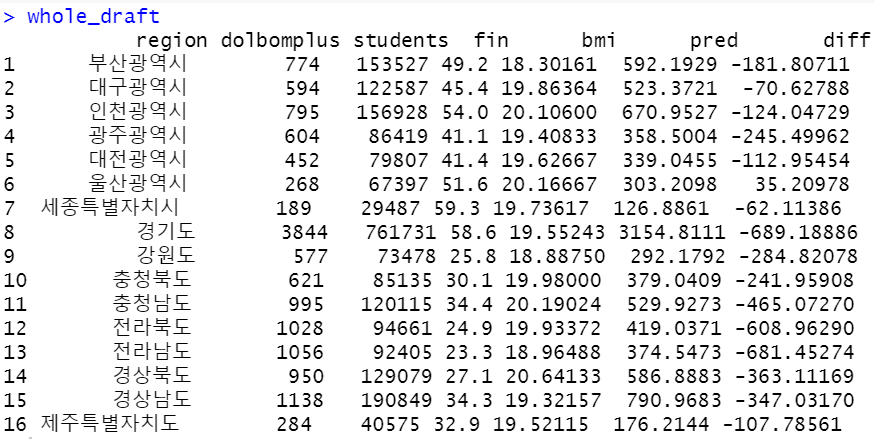


-한부모 제외 표준화 진행: 계수 절댓값 비교해보면 재정자립도가 다른 변수들 중 약하게 나타남. 하지만 stepwise 기법에서 재정자립도 선택하였기 때문에 stepwise 따라가기!

stepwise 결과

+: 변수 추가, - 변수 제거



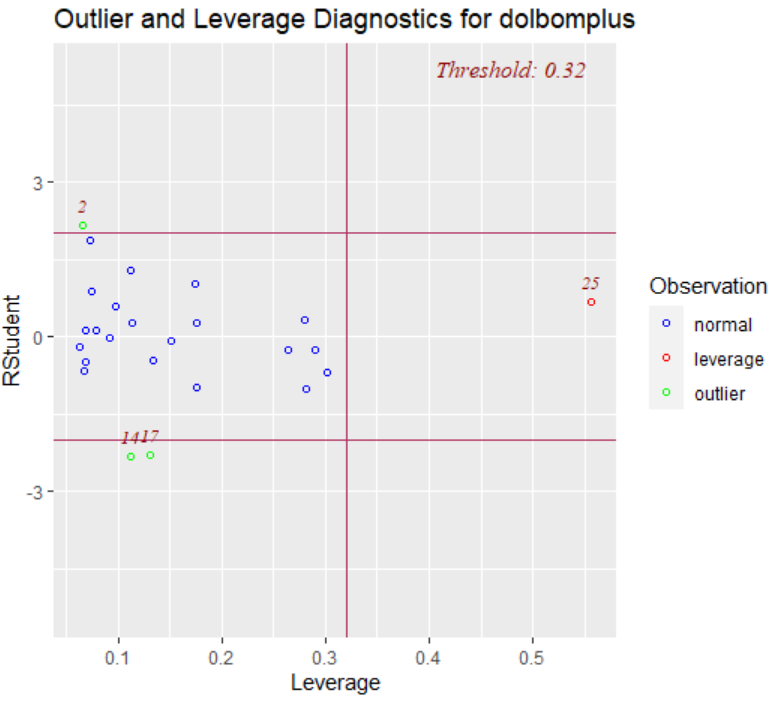


pred가 서울시 모델을 기준으로 해당 지역에 필요한 돌봄시설 숫자

diff가 pred에서 현재 돌봄시설개수(dombomplus)를 뺀 값

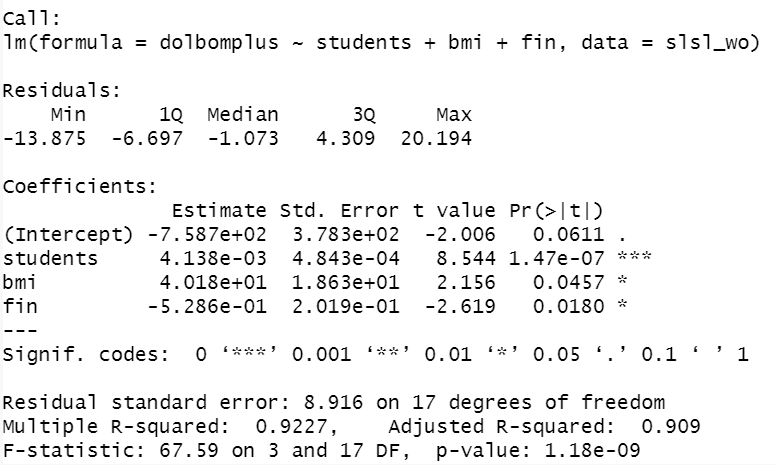
울산광역시 빼고는 다 부족

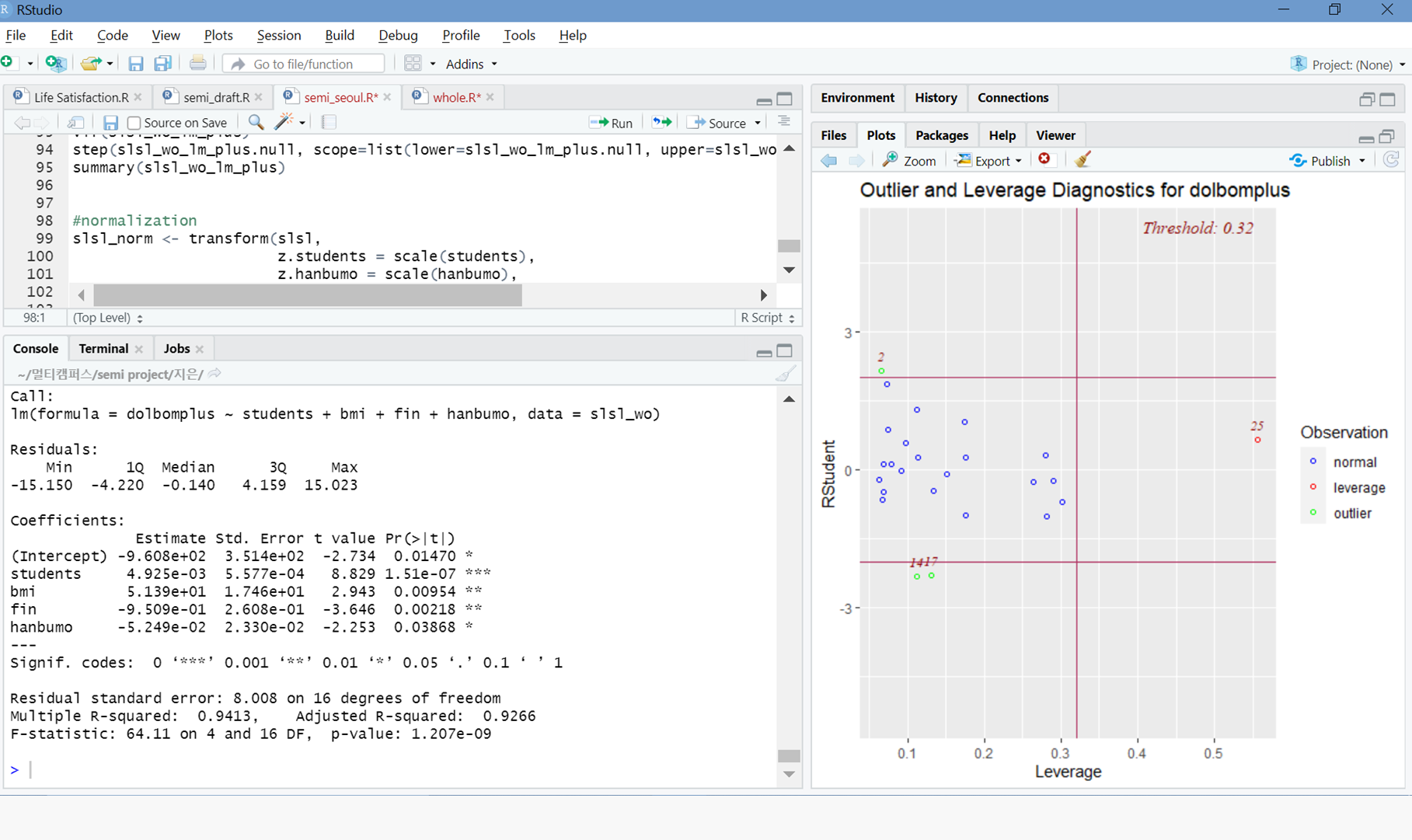
경기도를 예를 들면 현재 필요한 돌봄시설(학교교실+지역아동센터)이 총 3154개 정도인데 689개 정도가 부족하다



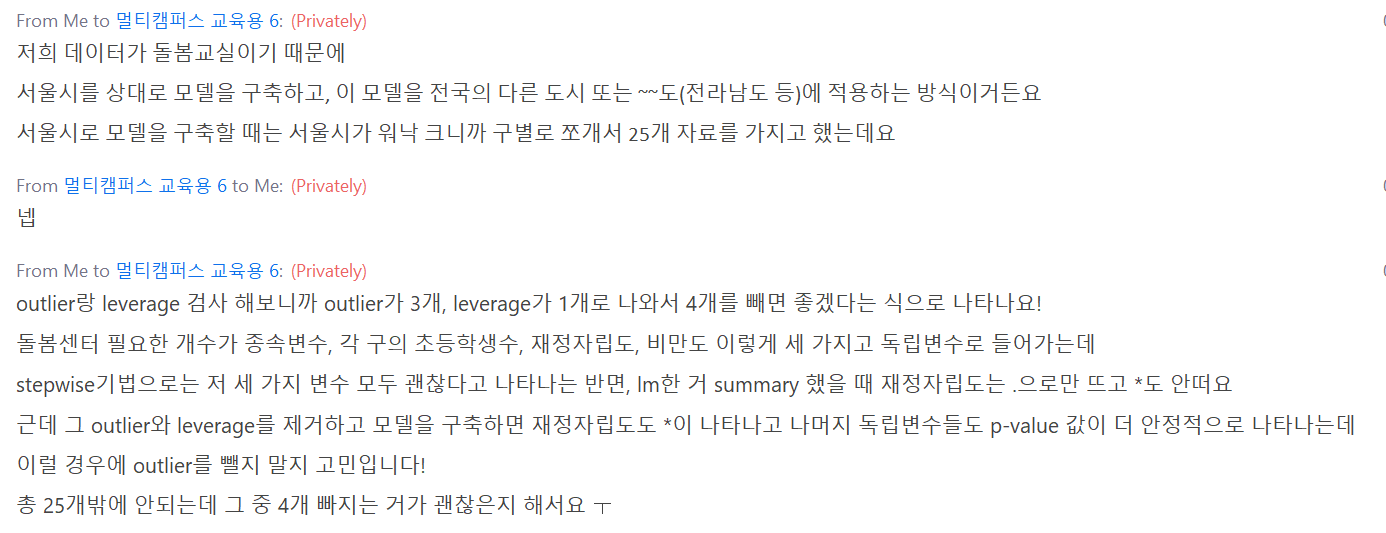
outlier랑 leverage: 범위에서 많이 벗어나는 표본 추출해줌 → 4개 범위에서 벗어남

-outlier랑 leverage 4개 제외 서울시 모델 다시 프로그래밍 결과: 학생, bmi, fin,한부모

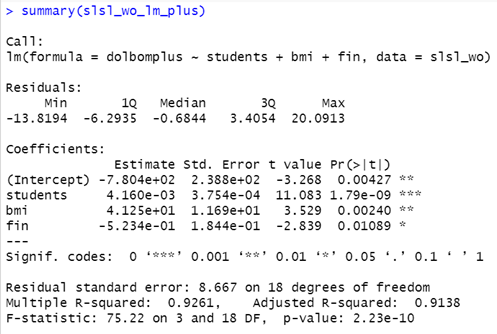




--범위 벗어나는 4개 제외해도 되는가?? --



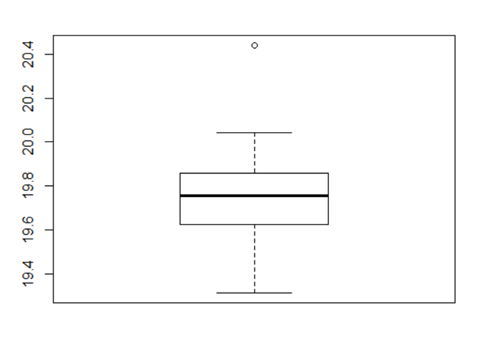
* outlier 만 제거하면 R squared value 0.9261



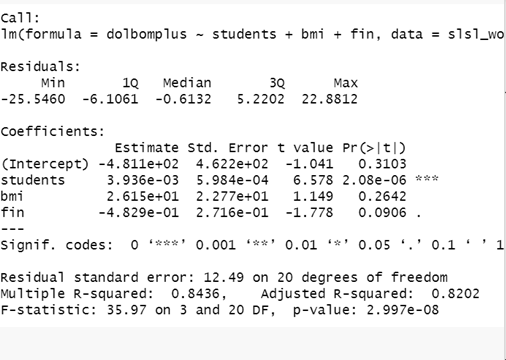
-간단하게 변수별로 boxplot을 그려서 그 내에서의 outlier가 있나 확인을 해보았는데요

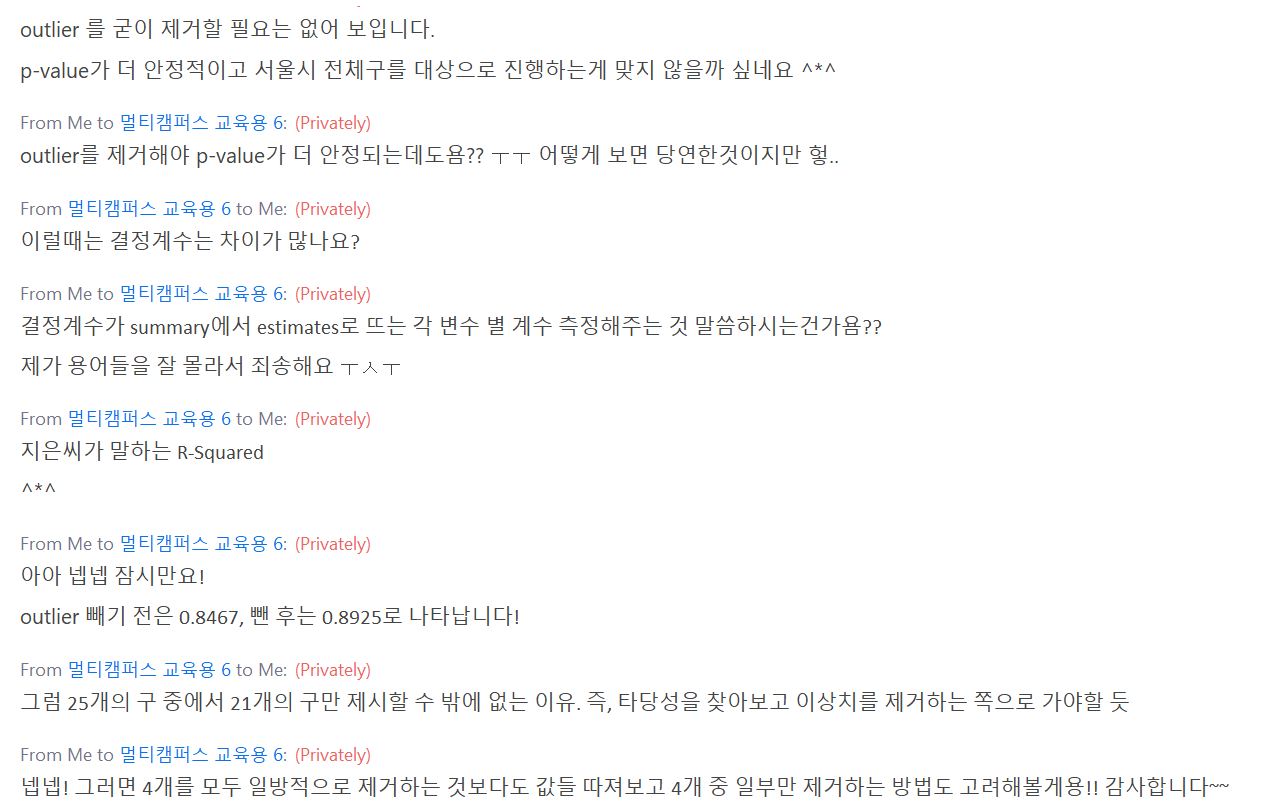
bmi에서만 outlier가 존재,

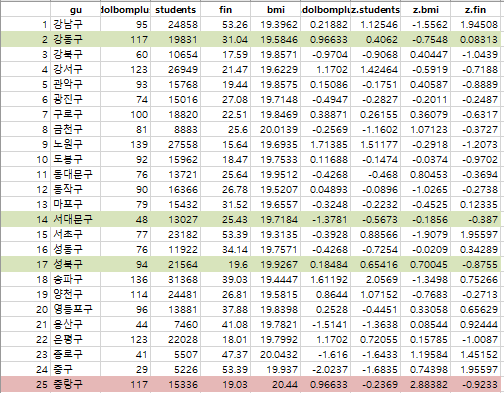
점: outlier도 아니라 leverage로 나오는 중랑구이거든요



outlier는 다시 다 넣은 채로 모델 구하는 걸 다시 해보았는데 no





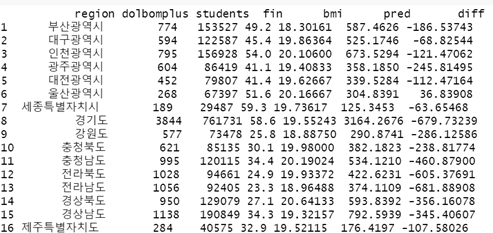


표준화를 하여 살펴보려고 해요 단위를 맞추면 좀 더 잘 보일까 해서요

z. = 표준화 진행

predict value(구축한 모델에 변수값을 넣어서 예측된 값)도 이용해서 difference(예측한 값과 실제값과의 차이) 계산한걸로 확인해봤는데 difference값이 가장 큰 세 개가 outlier

결과를 봐서 빼는 게 맞는 방향 같은데, 왜 빠져야 하는지 그 근거를 고민중이에요 왜 그 값이 이상치로 나타나는지



우선 outlier 세 개 모두 제외한 상태로 얻어낸 모델을 다른 지역에 적용해봤는데, 전부는 아니지만 대부분의 차이값이 줄었더라고요

outlier 3개를 제거하여 얻어낸 식은

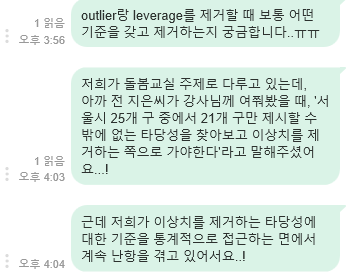
dolbomplus = -7.804e+02 + 4.160e-03 \* students + 4.125e+01 \* bmi + (-5.234e-01) \* fin

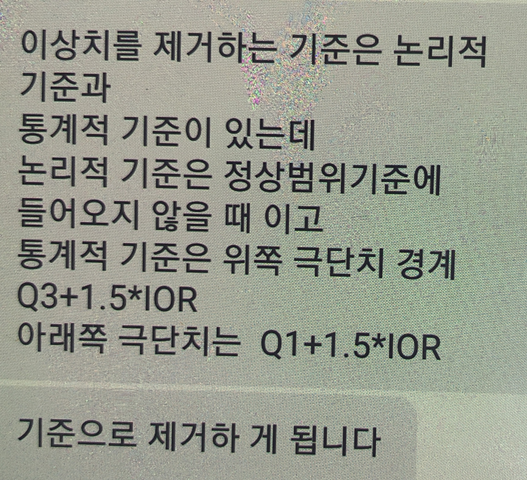
이렇게 돼요

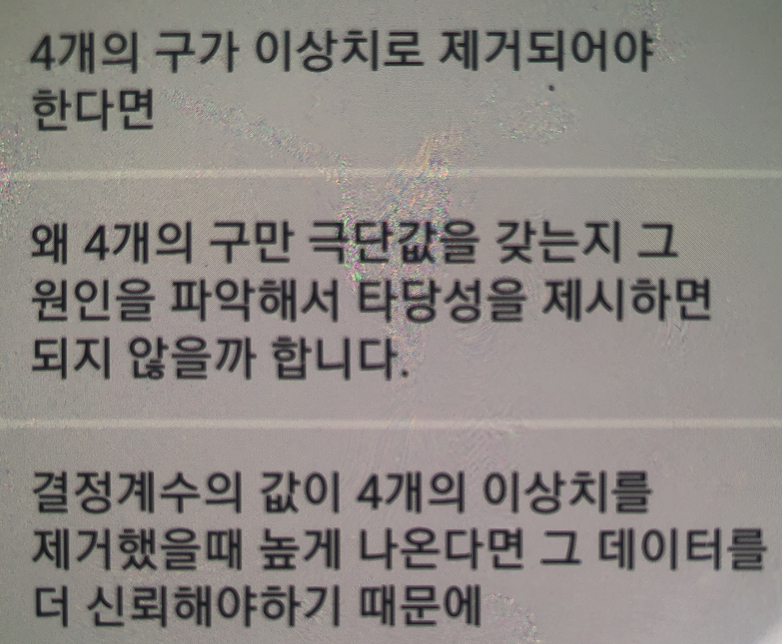
차이값이 경기도가 가장 컸는데 전라남도로 바꼇네요

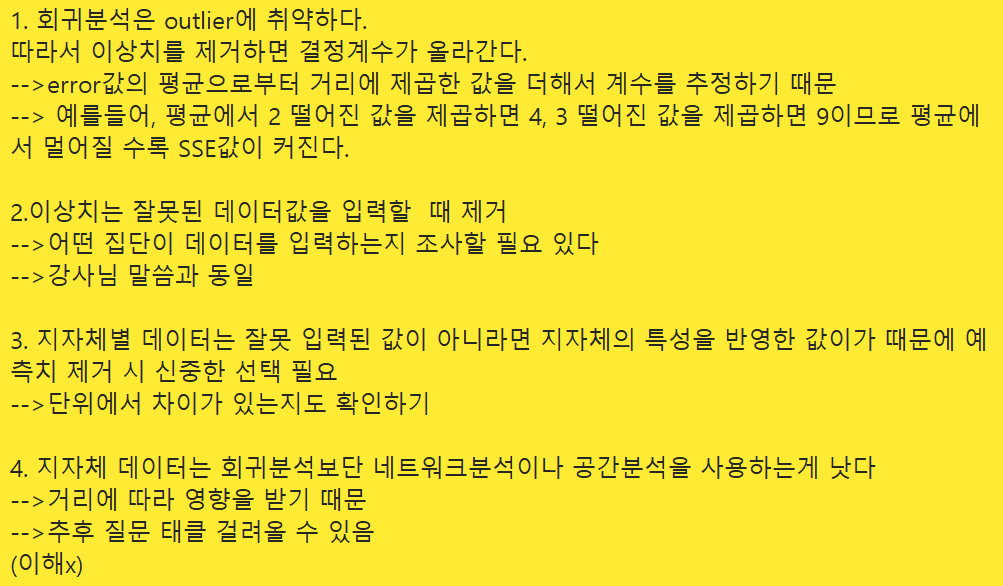
-----------------------------------------------------------------------------------------------------------

어떤 기준으로 outlier, leverage를 제거하는가…?





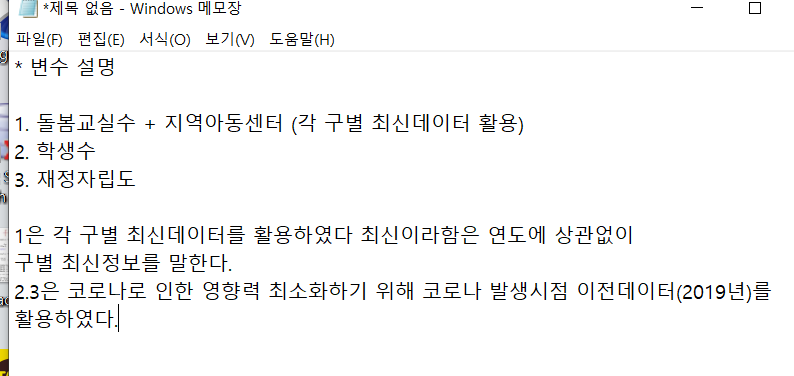




---------------------------------------------------------------------------------------------------------------------------9/2(수)

의료인력현황 관련한 변수는 어떨까요?? 코로나로 의료진 가정 돌봄이 시급하다고 하던데 정부가 의료진 가정에게 돌봄서비스 우선해서 제공하겠다고 하기도 한거 같아요

VIF 검사를 하면 지역내총생산과 재정자립도가 높게 나타나는 걸 볼 수 있거든용



돌봄교실수 = 학생수+ 비만+재정 + 교육

아니면 할말 많지만 1)데이터가 없거나 2)돌봄교실 수 측정 시 유의미한 값이 아닌 변수들

예를들어, 1)저소득, 한부모, 맞벌이,다문화,새터민 2)교육보조지원금,재정자립도 3) 의료인력

이런 값들이 향후 돌봄교실 수에 반영이 되어야 하는데 반영되지 않았다.

코로나 19 시대에는 위와 같은 변수를 고려하여 돌봄교실을 늘리고 서비스의 질적개선을 꾀해야 한다.

도서관은 도서관 많은 지역의 특징...? 도서관 설립 기준..? 등등을 통해서 도서관 개수와 그 지역구의 환경과 무슨 관련이 잇는지 알아보면 이야기할 때 도움될 것 같아요

한부모는 데이터가 잇지만 의미없게 나와서 사각지대로 얘기할 수 있고

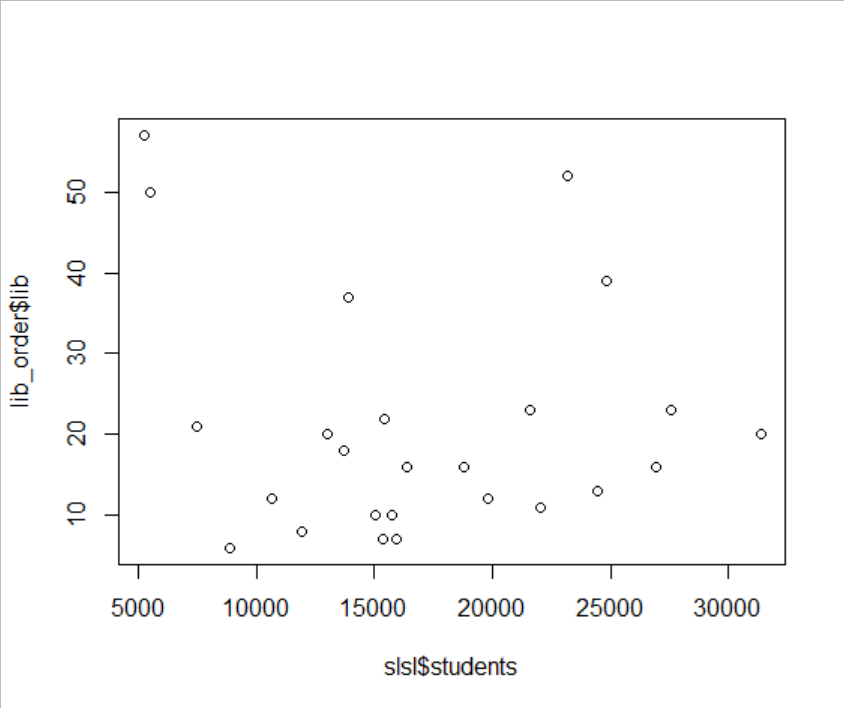
다문화 새터민은 데이터가 없엇지만 한부모와 비슷할 것 같다(이들의 사각지대 관련기사 잇다면 더 좋을 것 같아요)

그리고 코로나 시대를 언급하려면

코로나 시대가 갖는 특징이 무엇인지, 사용한 변수들이 그러한 특징에 어떠한 영향을 받는지 등등이 필요할 것 같아요

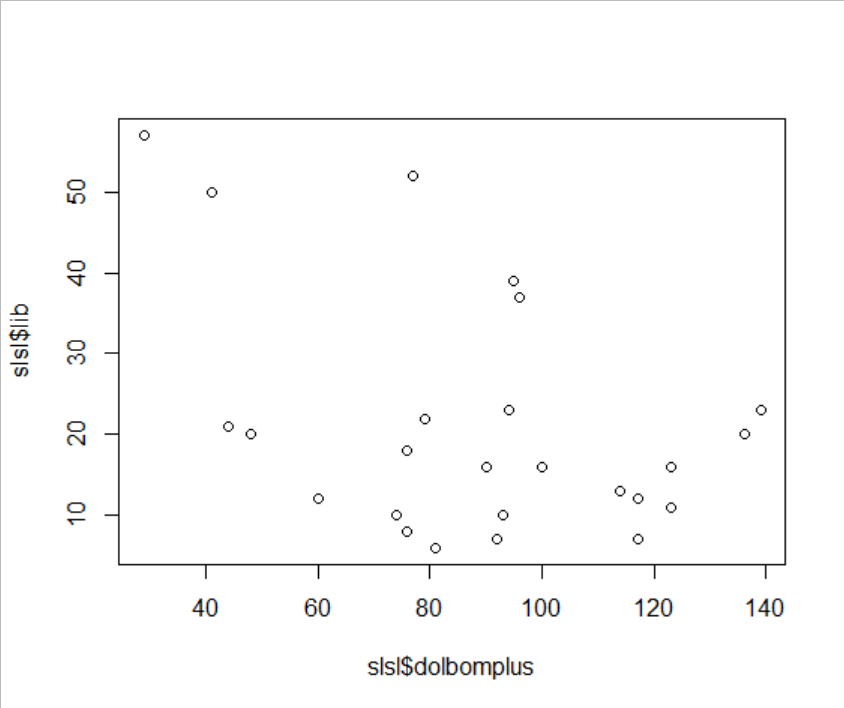
맞벌이는 정말 직관적으로 생각해도 의미있는 변수여서 꼭 반영하고 싶었는데 데이터가 존재하지 않앗다는 걸 가장 마지막(시사점 정도 언급하는...?) 그 부분에 짤막하게 언급하는 걸로 하는 게 좋을 것 같아요

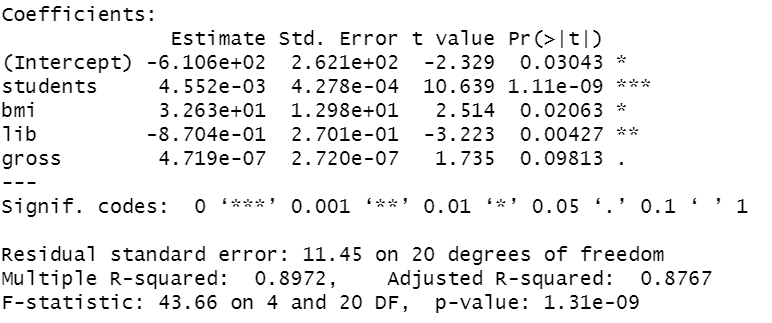
--학생수와 도서관의 상관성--



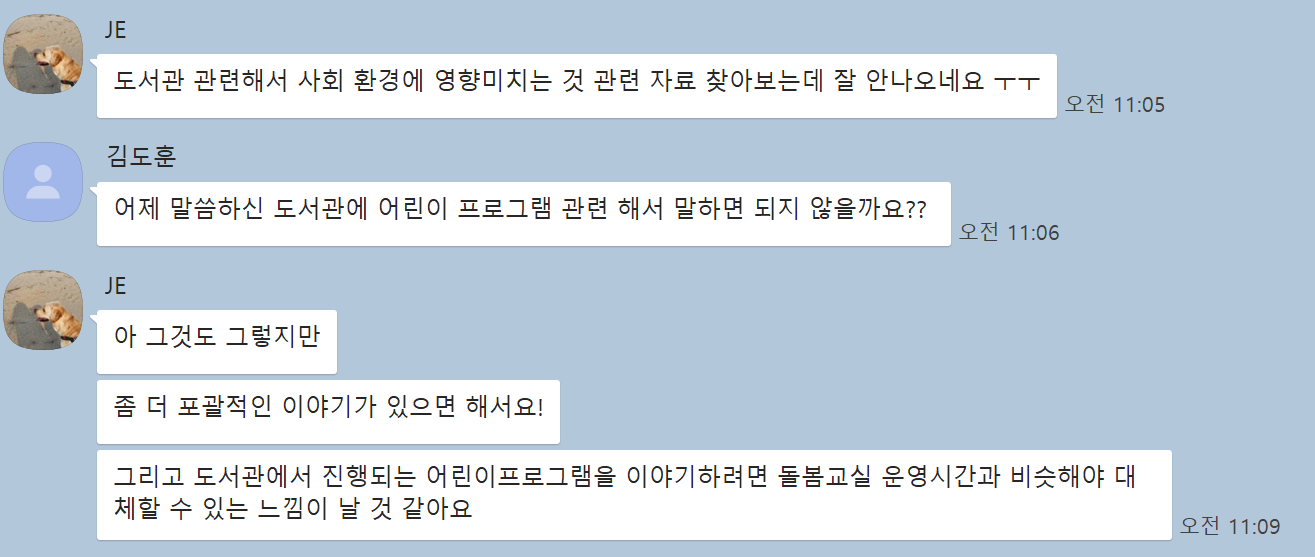
상관계수는 -0.1301356

dolbomplus랑 도서관은



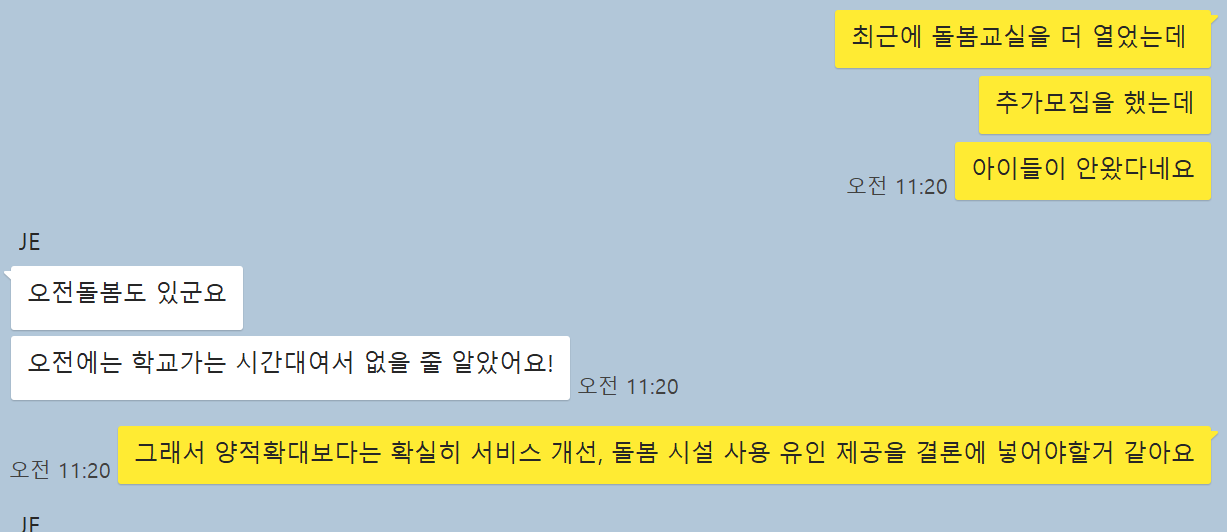


한부모,재정자립도,지역내총생산,구별살림규모도 하면 좋을 것 같아욤

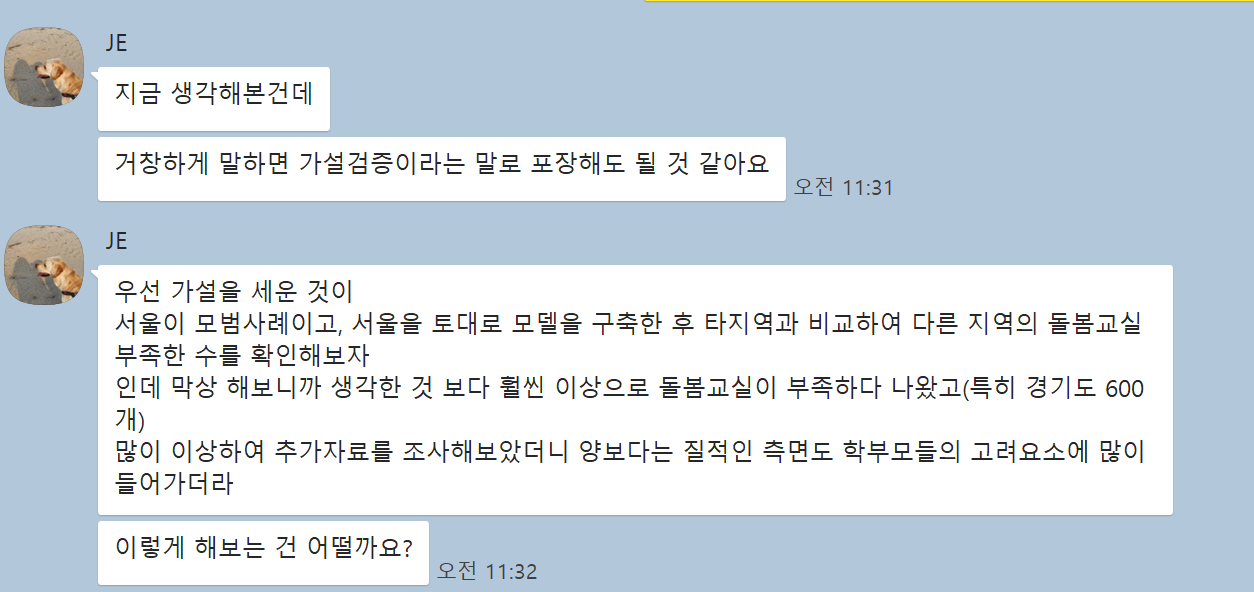


<http://www.segye.com/newsView/20200525503764>

<https://icare.seoul.go.kr/icare/dolbomSVC/dolbomSVC_7.jsp>

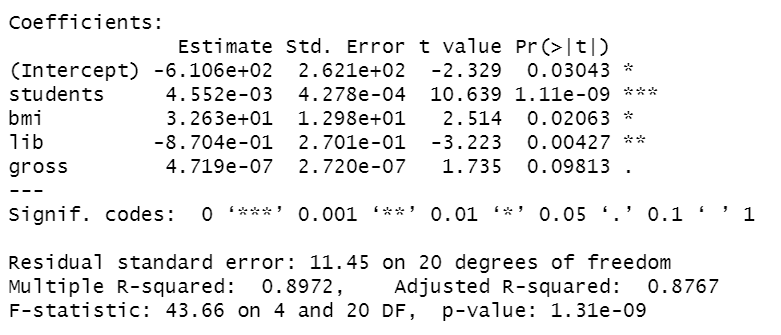


오늘 자료조사(들어간 변수 설명, 안들어가는 변수들 설명) 좀 부탁



뜻대로 되지 않음에 의문을 갖고 추가자료를 조사해보았다

9/3 최종



9/4(금)

