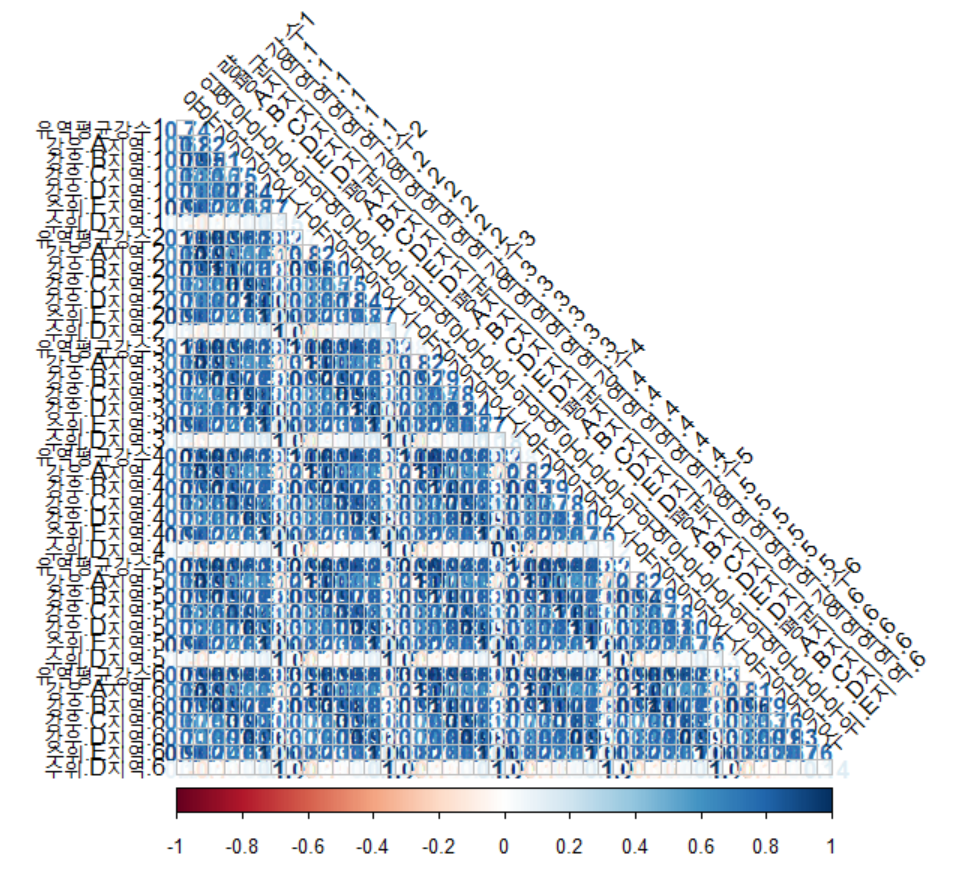
1. EDA

텍스트이(가) 표시된 사진

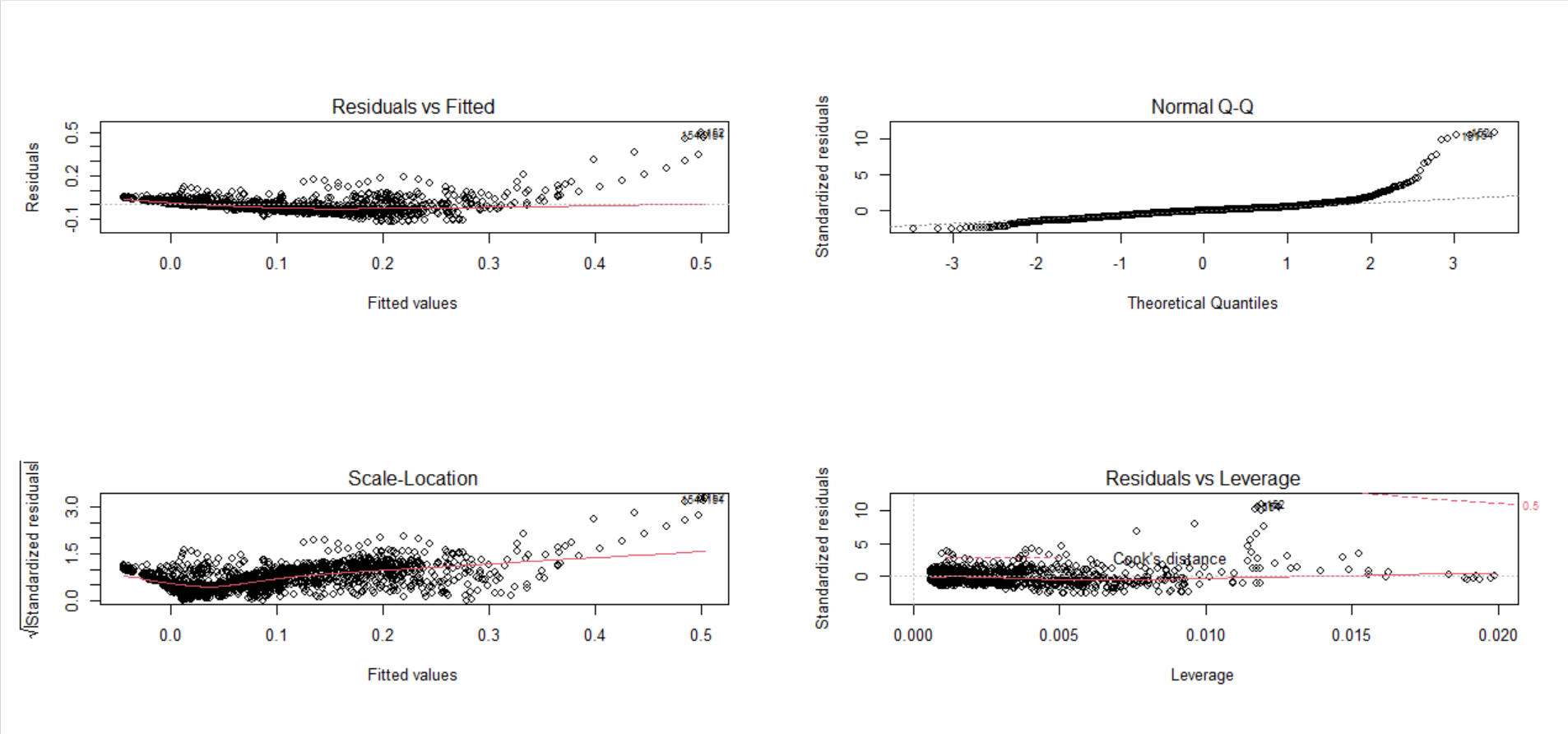
자동 생성된 설명

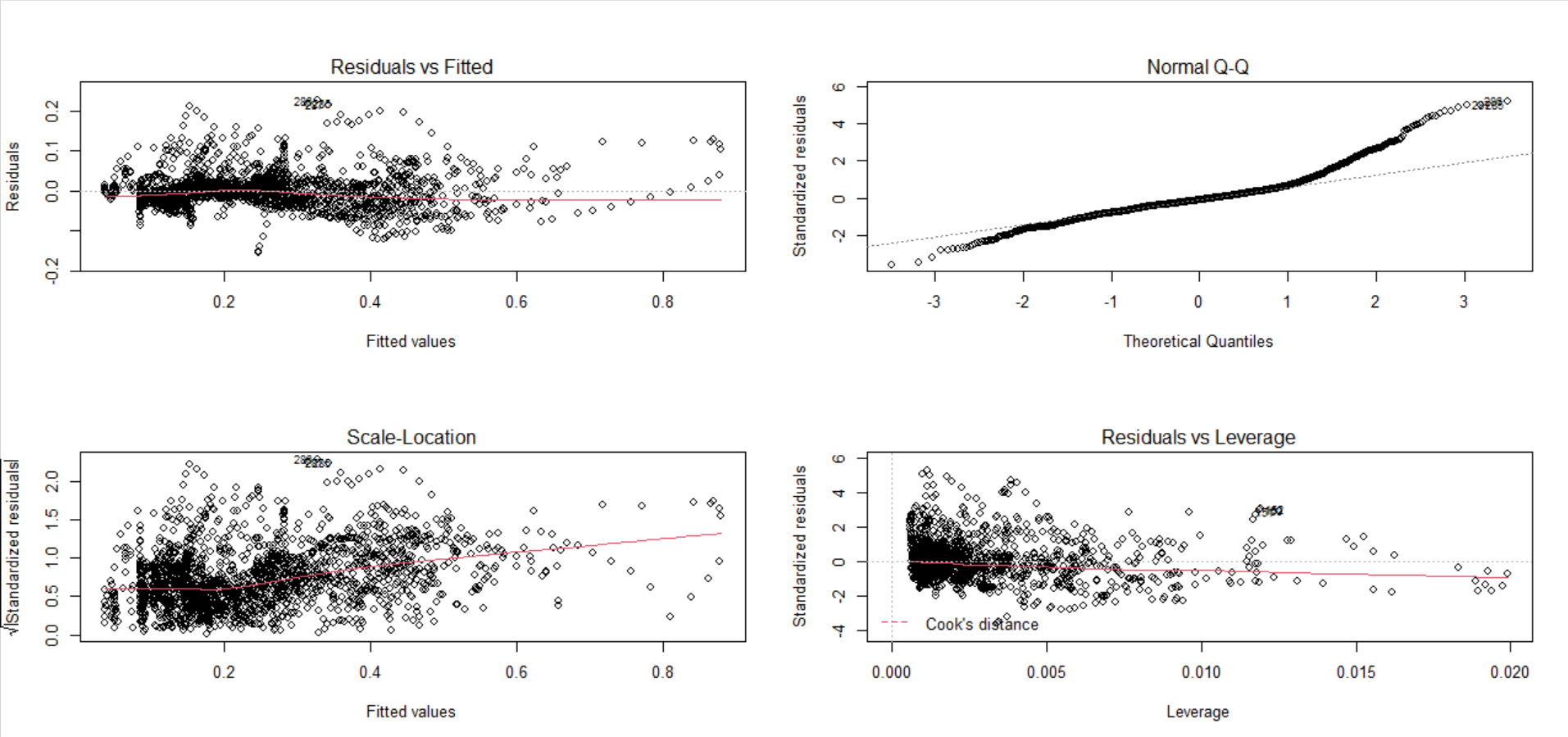
타겟변수는 연속형, 타겟변수 제외한 입력 피처들 42개, 총 2891행을 가지는 데이터셋, 고차원은 아님.

[시나리오1] 아웃라이어를 대치하지 않는다.



D제외하고는 상관계수 열라 높음-> 차원축소 필요(아웃라이어 보정하면 상관계수가 줄어들까..?? )

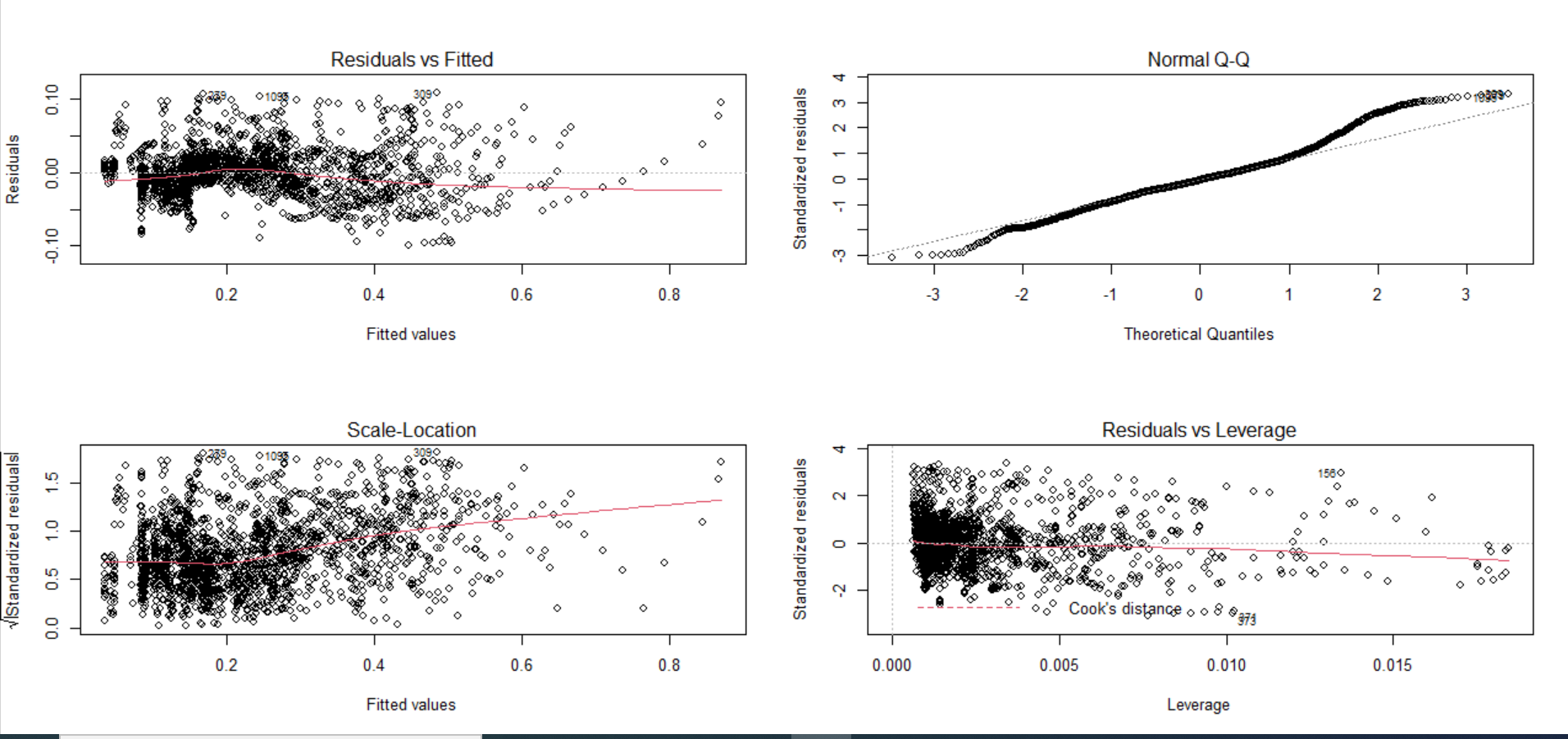
lm<-lm(target~., train) # 아무것도 하기 전 분산이 2차 형태를 띄고 있음 . 식을 통해서 검정하니



타겟에 제곱근 변환 함

① 등분산성 가정 이탈에 대한 처방

   오차항이 이분산성을 가지면 가중최소제곱추정법을 사용하거나 종속변수를 변수변환시켜서 최소제곱추정법을 적용한다. 변수변환의 방법으로는 제곱근변환, 로그변환, 역수변환, 이차식변환 , 지수변환등이 자주 사용된다.



포텐셜 아웃라이어 제거한 버젼

텍스트이(가) 표시된 사진

자동 생성된 설명

