**[두 번째 과제]**

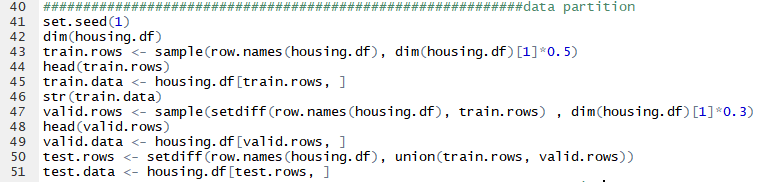
**IT응용시스템공학과**

**1494053**

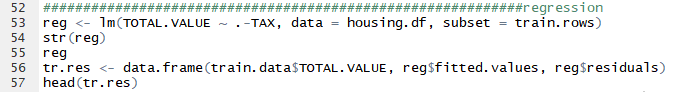
**김희택**

**WestRoxbury.csv를 훈련데이터(50%), 검증데이터(30%), 평가데이터(20%)로 분할하고, 훈련데이터를 이용해서 구한 선형회귀모형을 이용하여 훈련데이터, 그리고 검증데이터와 평가데이터에서 각 각 예측 값의 정확도를 비교하시오. (단 아래 조건에 주의하시오.)**

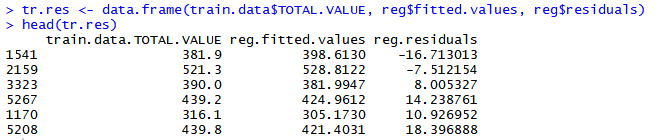
1. 목표변수는 TOTAL.VALUE 이며 선형회귀모형을 구할 때 TAX 변수를 사용하지 않는다.
2. 데이터를 분할할 때 set.seed(1)을 사용하시오.



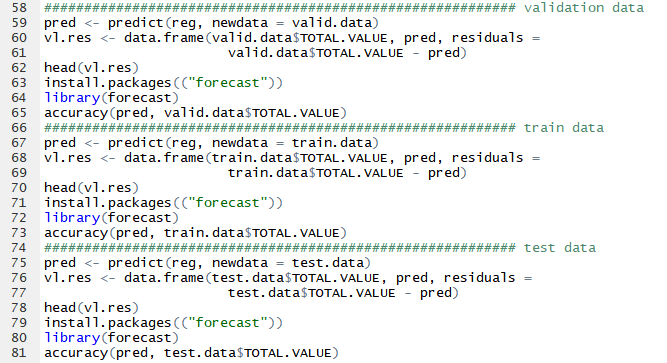
* 수업시간에는 훈련데이터 60%, 검증데이터 40%로 분할하였습니다. 이를 참고하여 두 번째 과제의 훈련데이터 50%, 검증데이터 30%, 평가데이터 20%로 분할하였습니다. 조건에 맞도록 데이터를 분할할 때 set.seed(1)을 사용하였습니다.



* 훈련데이터를 이용해서 선형회귀모형을 구하였습니다. 조건에 맞도록 목표변수는 TOTAL.VALUE 이며 TAX 변수를 사용하지 않았습니다.



* Partial Output 입니다.



* 선형회귀모형을 이용하여 훈련데이터, 그리고 검증데이터와 평가데이터에서 각 각 예측 값의 정확도를 구하였습니다.



* Output 입니다. 실제 값과 예측 값의 평균편차로 오차 메트릭을 확인해보면, 훈련데이터에서 예측 값의 정확도가, 오차 값이 가장 작으므로 모델예측이 검증데이터와 평가데이터보다 더 정확함을 알 수 있습니다.