

WestRoxbury.csv를 훈련데이터(50%), 검증데이터(30%), 평가데이터(20%)로 분할하고, 훈련 데이터를 이용해서 구한 선형회귀모형을 이용하여 훈련데이터, 검증데이터 그리고 평가데이터에서 각각의 예측값의 정확도를 비교하시오.

단, 아래 조건에 주의하시오.

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

```
setwd("c:/rdata")
housing.df <- read.csv("WestRoxbury.csv", header = TRUE)

set.seed(1)
#훈련데이터#
train.rows <- sample(row.names(housing.df), dim(housing.df)[1]*0.5)

#검증데이터#
valid.rows <- sample(setdiff(rownames(housing.df), train.rows), dim(housing.df)[1]*0.3)

#평가데이터#
test.rows <- setdiff(rownames(housing.df), union(train.rows, valid.rows))

#create the three data frame#
train.data <- housing.df[train.rows, ]
valid.data <- housing.df[valid.rows, ]
test.data <- housing.df[test.rows, ]

#linear model#
reg <- lm(TOTAL.VALUE ~ .-TAX, data = housing.df, subset = train.rows)
pred.valid <- predict(reg, newdata = valid.data)
pred.test <- predict(reg, newdata = test.data)

> accuracy(reg$fitted.values, train.data$TOTAL.VALUE)
              ME      RMSE      MAE      MPE      MAPE
Test set -6.662869e-16 42.33657 32.07316 -1.084396 8.443679
> accuracy(pred.valid, valid.data$TOTAL.VALUE)
              ME      RMSE      MAE      MPE      MAPE
Test set 1.080142 42.96173 31.91816 -0.653073 8.14583
> accuracy(pred.test, test.data$TOTAL.VALUE)
              ME      RMSE      MAE      MPE      MAPE
Test set 1.177264 45.04838 33.55389 -0.9799544 8.72732
```