Linear Regression Case Study: Boston Housing Prices

Background:

The data has 506 cases where each case is a location in Boston. The “median housing price” is a target variable. The data has 3 variables which are described in the table below

The objective is to identify significant factors affecting housing prices.

Import data and display first 6 rows

```
hp<-read.csv("Housing Prices.csv",header=T)
head(hp,6)
```

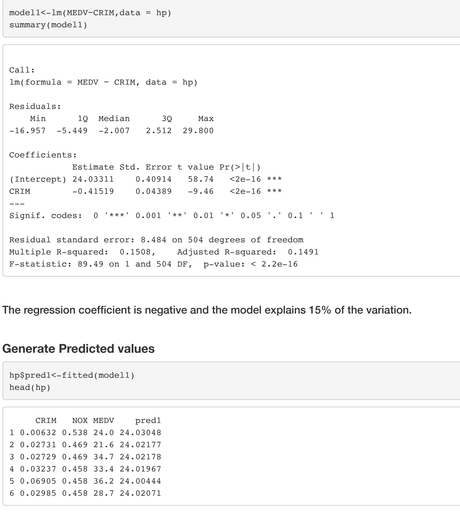
```
     CRIM   NOX MEDV
1 0.00632 0.538 24.0
2 0.02731 0.469 21.6
3 0.02729 0.469 34.7
4 0.03237 0.458 33.4
5 0.06905 0.458 36.2
6 0.02985 0.458 28.7
```

Data Description

Column name	Column description
CRIM	Per Capita Crime Rate by town
NOX	Nitric Oxides concentration (parts per 10 million)
MEDV	Median value of owner-occupied homes in $1000's

Scatter plot matrix

```
library(GGally)
ggpairs(hp, title = "Scatter Plot Matrix")
```



Simple Linear Regression: MEDV vs CRIM

```
model1<-lm(MEDV~CRIM,data = hp)
summary(model1)
```

```

Call:
lm(formula = MEDV ~ CRIM, data = hp)

Residuals:
    Min      1Q  Median      3Q     Max
-16.957  -5.449  -2.007   2.512  29.800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 24.03311    0.40914   58.74   <2e-16 ***
CRIM        -0.41519    0.04389   -9.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.484 on 504 degrees of freedom
Multiple R-squared:  0.1508,    Adjusted R-squared:  0.1491
F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

The regression coefficient is negative and the model explains 15% of the variation.

Generate Predicted values

```
hp$pred1<-fitted(model1)
head(hp)
```

```
     CRIM   NOX MEDV    pred1
1 0.00632 0.538 24.0 24.03048
2 0.02731 0.469 21.6 24.02177
3 0.02729 0.469 34.7 24.02178
4 0.03237 0.458 33.4 24.01967
5 0.06905 0.458 36.2 24.00444
6 0.02985 0.458 28.7 24.02071
```

Simple Linear Regression: MEDV vs NOX

```
model2<-lm(MEDV~NOX,data = hp)
summary(model2)
```

```

Call:
lm(formula = MEDV ~ NOX, data = hp)

Residuals:
    Min      1Q  Median      3Q     Max
-13.691  -5.121  -2.161   2.959  31.310

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   41.346      1.811   22.83   <2e-16 ***
NOX          -33.916      3.196  -10.61   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.323 on 504 degrees of freedom
Multiple R-squared:  0.1826,    Adjusted R-squared:  0.181
F-statistic: 112.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

The regression coefficient is negative and the model explains 18.26% of the variation.

Generate Predicted values

```
hp$pred2<-fitted(model2)
head(hp)
```

```
     CRIM   NOX MEDV    pred1    pred2
1 0.00632 0.538 24.0 24.03048 23.09904
2 0.02731 0.469 21.6 24.02177 25.43924
3 0.02729 0.469 34.7 24.02178 25.43924
4 0.03237 0.458 33.4 24.01967 25.81232
5 0.06905 0.458 36.2 24.00444 25.81232
6 0.02985 0.458 28.7 24.02071 25.81232
```

Multiple Linear Regression

```
model4<-lm(MEDV~CRIM+NOX,data=hp)
summary(model4)
```

```

Call:
lm(formula = MEDV ~ CRIM + NOX, data = hp)

Residuals:
    Min      1Q  Median      3Q     Max
-15.080  -5.005  -2.207   2.539  31.611

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.62994    1.86339  20.194  < 2e-16 ***
CRIM         -0.27084    0.04596  -5.893 6.95e-09 ***
NOX         -25.45261    3.41148  -7.461 3.80e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.058 on 503 degrees of freedom
Multiple R-squared:  0.2354,    Adjusted R-squared:  0.2324
F-statistic: 77.43 on 2 and 503 DF,  p-value: < 2.2e-16
```

The regression coefficients are negative and the model explains 23.54% of the variation.

Generate Predicted values

```
hp$pred4<-fitted(model4)
head(hp)
```

```
     CRIM   NOX MEDV    pred1    pred2    pred4
1 0.00632 0.538 24.0 24.03048 23.09904 23.93473
2 0.02731 0.469 21.6 24.02177 25.43924 25.68527
3 0.02729 0.469 34.7 24.02178 25.43924 25.68528
4 0.03237 0.458 33.4 24.01967 25.81232 25.96388
5 0.06905 0.458 36.2 24.00444 25.81232 25.95394
6 0.02985 0.458 28.7 24.02071 25.81232 25.96456
```