Homework 1 - IND2627

Daniel L Santos

2025-03-23

Exercise 1

```
#Necessary packages for this exercise
library(tidyverse)
library(MASS)
```

The Boston dataset is a dataset derived from information collected by the U.S Census Service concerning housing in the area of Boston Mass. It is composed by

The Table 1 shows the first six observations from the dataset.

Table 1: Boston dataset

crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
0.00632	18	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
0.02731	0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
0.02729	0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
0.03237	0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
0.06905	0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
0.02985	0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

Fitting the multiple linear regression

```
boston_regression_model <- lm(medv ~ lstat + age, Boston)
boston_regression_model
```

```
Call:
lm(formula = medv ~ lstat + age, data = Boston)

Coefficients:
(Intercept)        lstat          age
   33.22276     -1.03207      0.03454
```

As we can observe, the estimated model function $\hat{Y} = \hat{f}(X)$ is:

$$\hat{medv} = -1.03207 * lstat + 0.03454 * age + 33.22276.$$

The estimated coefficients are $\hat{\beta}_0 = 33.22276$, $\hat{\beta}_1 = -1.03207$ and $\hat{\beta}_2 = 0.03454$. Notice that the coefficient $\hat{\beta}_1$ is negative, which means that, for a fixed value of *age*, a one-unit increase in *lstat* leads to a decrease in *medv* by 1.03207 units. On the other hand, for a fixed value of *lstat*, a one-unit increase in *age* leads to an increase in *medv* by 0.03454 units. Finally, when both *age* and *lstat* are zero, the average value of *medv* will be $\hat{\beta}_0 = 33.22276$.

```r
summary(boston_regression_model)
```

```
Call:
lm(formula = medv ~ lstat + age, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-15.981  -3.978  -1.283   1.968  23.158

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.22276    0.73085  45.458  < 2e-16 ***
lstat       -1.03207    0.04819 -21.416  < 2e-16 ***
age          0.03454    0.01223   2.826  0.00491 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.173 on 503 degrees of freedom
Multiple R-squared:  0.5513,    Adjusted R-squared:  0.5495
F-statistic:   309 on 2 and 503 DF,  p-value: < 2.2e-16
```

```r
confint(boston_regression_model)
```

```
                  2.5 %      97.5 %
(Intercept) 31.78687150 34.65864956
lstat       -1.12674848 -0.93738865
age          0.01052507  0.05856361
```