Homework 1

Kirill Lesnykh

January 19, 2019

Importing the dataset

```
df <- read.csv("carprice.csv")
df
```

```
##    Car Age Miles Price
## 1    1   5    57    85
## 2    2   4    40   103
## 3    3   6    77    70
## 4    4   5    60    82
## 5    5   5    49    89
## 6    6   5    47    98
## 7    7   6    58    66
## 8    8   6    39    95
## 9    9   2     8   169
## 10  10   7    69    70
## 11  11   7    89    48
```

Creating the vectors "x" and "y"

```
x <- c(df$Age)
x
```

```
##  [1] 5 4 6 5 5 5 6 6 2 7 7
```

```
y <- c(df$Price)
y
```

```
##  [1]  85 103  70  82  89  98  66  95 169  70  48
```

Conducting a linear regression

```
lm1 <- lm(y ~ x)
summary(lm1)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.162  -8.531  -5.162   8.946  21.099
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   195.47      15.24  12.826 4.36e-07 ***
## x             -20.26       2.80  -7.237 4.88e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.58 on 9 degrees of freedom
## Multiple R-squared:  0.8534, Adjusted R-squared:  0.8371
## F-statistic: 52.38 on 1 and 9 DF,  p-value: 4.882e-05
```

Conducting ANOVA test

```
ANOVA <- aov(y ~ x, df)
summary(ANOVA)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## x            1   8285    8285   52.38 4.88e-05 ***
## Residuals    9   1424     158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Due to the results of the ANOVA test, on the confidence level 95% the P-value is significant. So, there is significant difference betwen the Age and the Price of cars.

Runing the given code (constructing the Sum Sq manually)

```
n <- 11

X <- cbind(1,x)

H <- X%*%solve(t(X)%*%X)%*%t(X)

J <- matrix(1,n,n)

In <- diag(n)

SStotal <- t(y)%*%(In-1/n*J)%*%y

SSreg <-  t(y)%*%(H-1/n*J)%*%y

SSres <- t(y)%*%(In-H)%*%y

SSreg; SSreg
```

```
##          [,1]
## [1,] 8285.014
```

```
##          [,1]
## [1,] 8285.014
```

Based on the comparing with the results of the ANOVA test, the code that just was run, it calculates the Sum of the Squares of the linear regression model