Regression.R

User02

Wed Mar 21 12:52:15 2018

```
setwd("C:/ ")
Housing=read.csv("HousingData.csv", header=TRUE)
Housing=Housing[,-1]
head(Housing)
```

```
##   SellingPrice000s HouseSize00Sqft NumberofBathrms NumberofBedrms
## 1              290              21               2              4
## 2               95              11               1              2
## 3              170              19               2              3
## 4              375              38               4              5
## 5              350              24               3              4
## 6              125              10               2              2
##   GarageSize
## 1          2
## 2          0
## 3          2
## 4          3
## 5          2
## 6          0
```

```
summary(Housing)
```

```
##  SellingPrice000s HouseSize00Sqft NumberofBathrms NumberofBedrms
##  Min.   : 95.0    Min.   :10.00   Min.   :1.00    Min.   :2.00
##  1st Qu.:185.0    1st Qu.:19.00   1st Qu.:2.00    1st Qu.:4.00
##  Median :275.0    Median :24.00   Median :3.00    Median :4.00
##  Mean   :261.8    Mean   :23.96   Mean   :2.76    Mean   :3.92
##  3rd Qu.:340.0    3rd Qu.:29.00   3rd Qu.:4.00    3rd Qu.:4.00
##  Max.   :430.0    Max.   :38.00   Max.   :4.00    Max.   :5.00
##    GarageSize
##  Min.   :0
##  1st Qu.:2
##  Median :2
##  Mean   :2
##  3rd Qu.:2
##  Max.   :3
```

```
str(Housing)
```

```
## 'data.frame':    25 obs. of  5 variables:
##  $ SellingPrice000s: int  290 95 170 375 350 125 310 275 340 215 ...
##  $ HouseSize00Sqft : int  21 11 19 38 24 10 31 25 27 22 ...
##  $ NumberofBathrms : int  2 1 2 4 3 2 4 2 3 3 ...
```

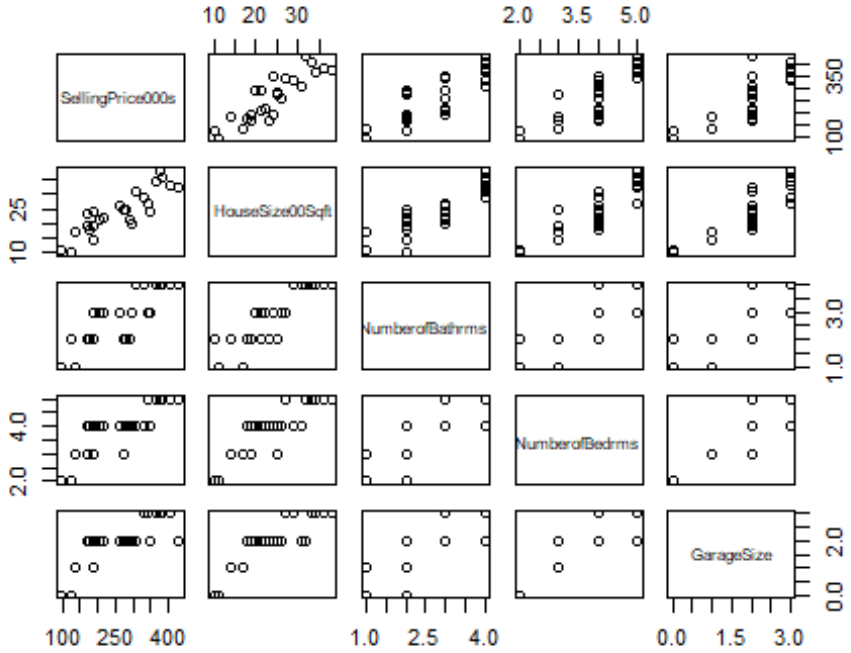
```
##  $ NumberofBedrms  : int  4 2 3 5 4 2 4 3 5 4 ...
##  $ GarageSize      : int  2 0 2 3 2 0 2 2 3 2 ...
```

```
cor(Housing)
```

```
##                  SellingPrice000s HouseSize00Sqft NumberofBathrms
## SellingPrice000s        1.0000000       0.8772285       0.8320072
## HouseSize00Sqft         0.8772285       1.0000000       0.8527317
## NumberofBathrms         0.8320072       0.8527317       1.0000000
## NumberofBedrms          0.8182372       0.8554267       0.7736148
## GarageSize              0.7686695       0.8555314       0.7368829
##                  NumberofBedrms GarageSize
## SellingPrice000s      0.8182372  0.7686695
## HouseSize00Sqft       0.8554267  0.8555314
## NumberofBathrms       0.7736148  0.7368829
## NumberofBedrms        1.0000000  0.8878382
## GarageSize            0.8878382  1.0000000
```

```
pairs(Housing)
```

```
library(sp)
```




```
library(raster)
library(usdm)
```

```
## Warning: package 'usdm' was built under R version 3.4.4
```

```
vifstep(Housing[,-1], th=5)
```

```
## 2 variables from the 4 input variables have collinearity problem:
##
## HouseSize00Sqft NumberofBedrms
##
## After excluding the collinear variables, the linear correlation
coefficients ranges between:
## min correlation ( GarageSize ~ NumberofBathrms ):  0.7368829
## max correlation ( GarageSize ~ NumberofBathrms ):  0.7368829
##
## ---------- VIFs of the remained variables --------
##         Variables      VIF
## 1 NumberofBathrms 2.188167
## 2      GarageSize 2.188167
```

```
fit1=lm(SellingPrice000s~NumberofBathrms + GarageSize, data=Housing)
summary(fit1)
```

```
##
## Call:
## lm(formula = SellingPrice000s ~ NumberofBathrms + GarageSize,
##     data = Housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -85.553 -42.716   2.284  24.588  97.144
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        23.92      31.43   0.761  0.45469
## NumberofBathrms    57.30      15.70   3.651  0.00141 **
## GarageSize         39.86      18.64   2.138  0.04384 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.4 on 22 degrees of freedom
## Multiple R-squared:  0.7452, Adjusted R-squared:  0.722
## F-statistic: 32.17 on 2 and 22 DF,  p-value: 2.939e-07
```

```
reduced=lm(SellingPrice000s~NumberofBathrms, data=Housing)
full=lm(SellingPrice000s~NumberofBedrms + HouseSize00Sqft, data=Housing)
anova(reduced,full)
```

```
## Analysis of Variance Table
##
## Model 1: SellingPrice000s ~ NumberofBathrms
## Model 2: SellingPrice000s ~ NumberofBedrms + HouseSize00Sqft
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1     23 67506
## 2     22 46790  1     20716 9.7406 0.004975 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
predict(fit1,data.frame(NumberofBathrms=3,GarageSize=3),
        interval="confidence", level=0.95)
```

```
##        fit     lwr      upr
## 1 315.4122 276.077 354.7474
```

```r
predict(fit1,data.frame(NumberofBathrms=3,GarageSize=3),
        interval="prediction", level=0.95)
```

```
##        fit      lwr      upr
## 1 315.4122 203.7277 427.0968
```