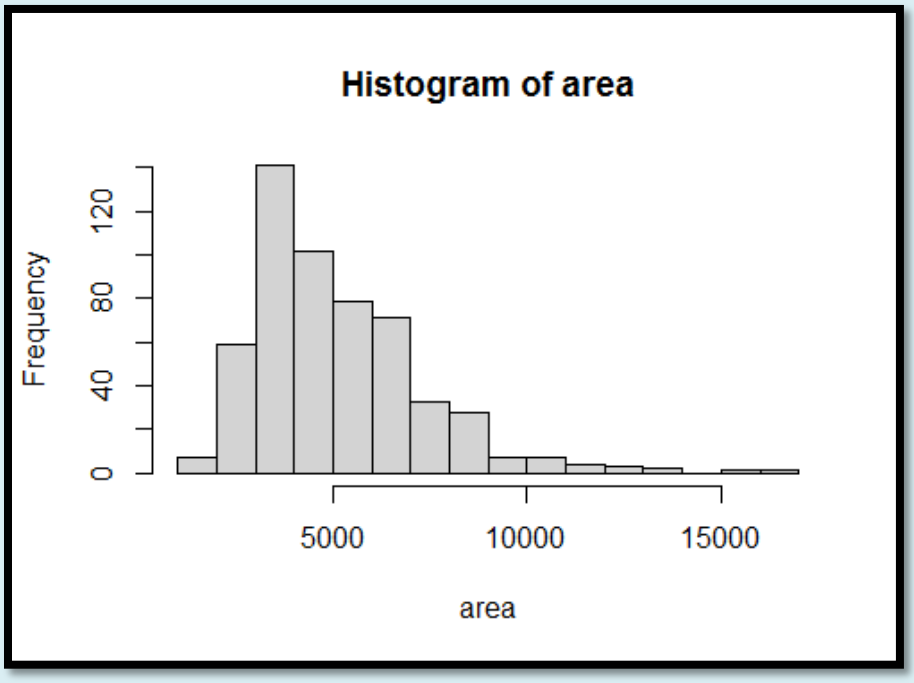
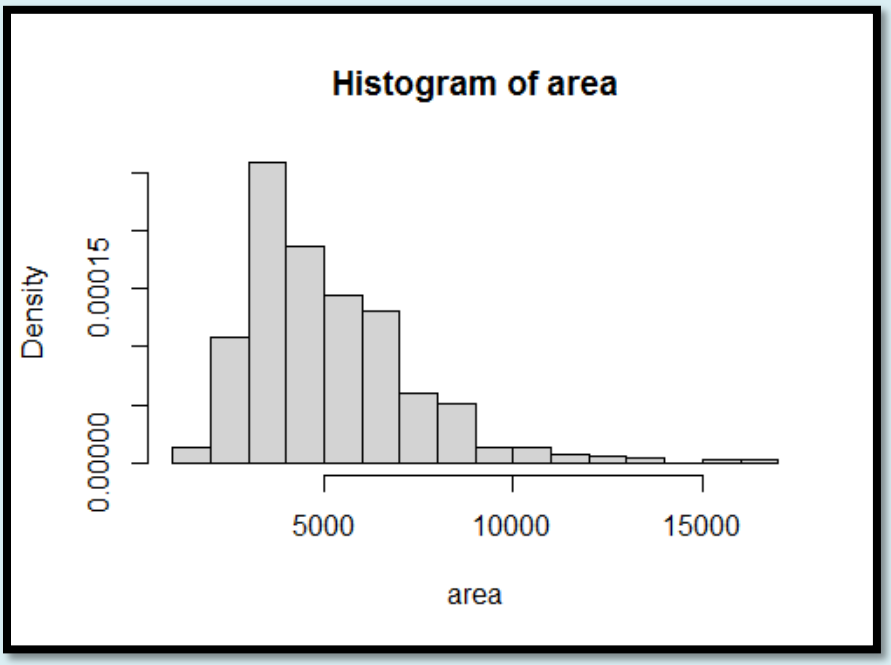
# Linear Regression using R- Programming

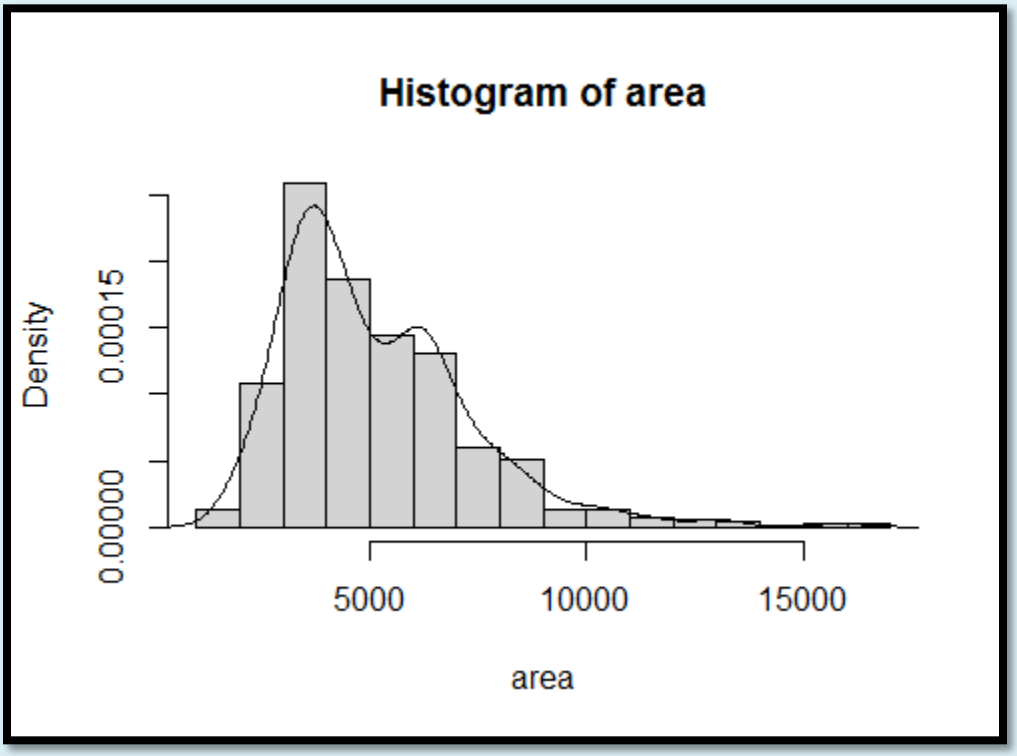
```
> mydata=read.csv(file.choose(), header = T)
> View(mydata)
> attach(mydata)
> #check the dimension of the data set
> dim(mydata)
[1] 545  13
> #to check the variable names
> names(mydata)
 [1] "price"          "area"           "bedrooms"       "bathrooms"      "stories"
 [6] "mainroad"       "guestroom"      "basement"       "hotwaterheating"  "airconditioning"
[11] "parking"        "prefarea"       "furnishingstatus"
> #to check normality
> hist(area)
```



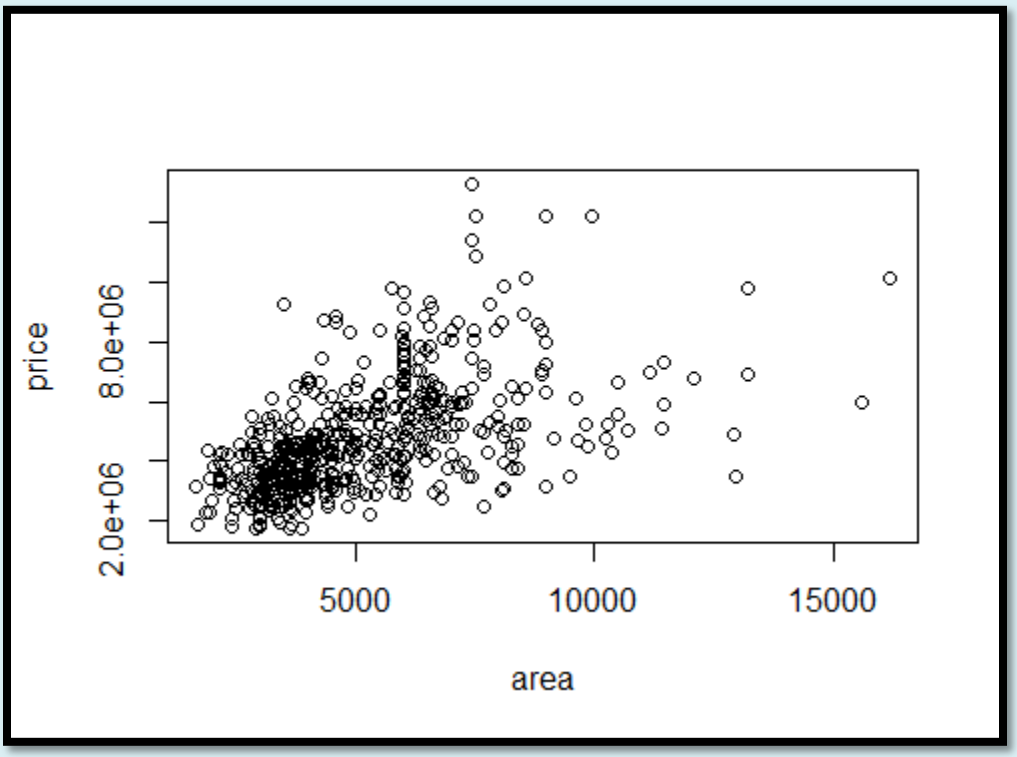

```
> hist(area, freq = F)
```



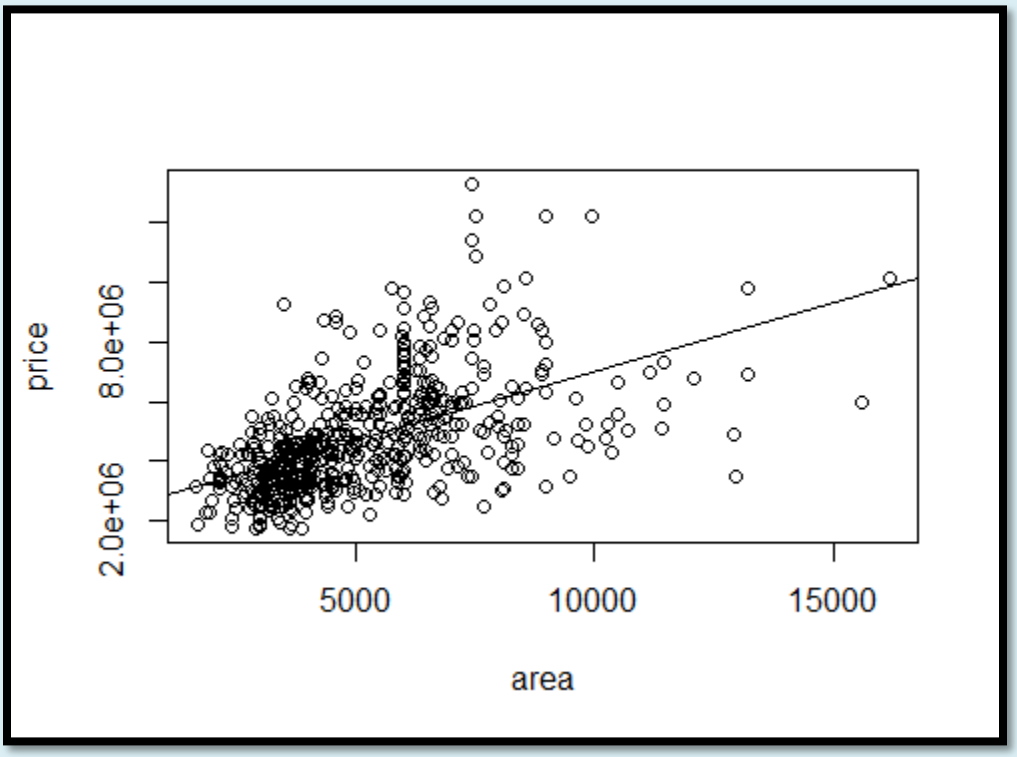

**> lines(density(area))**




**> plot(price~area)**

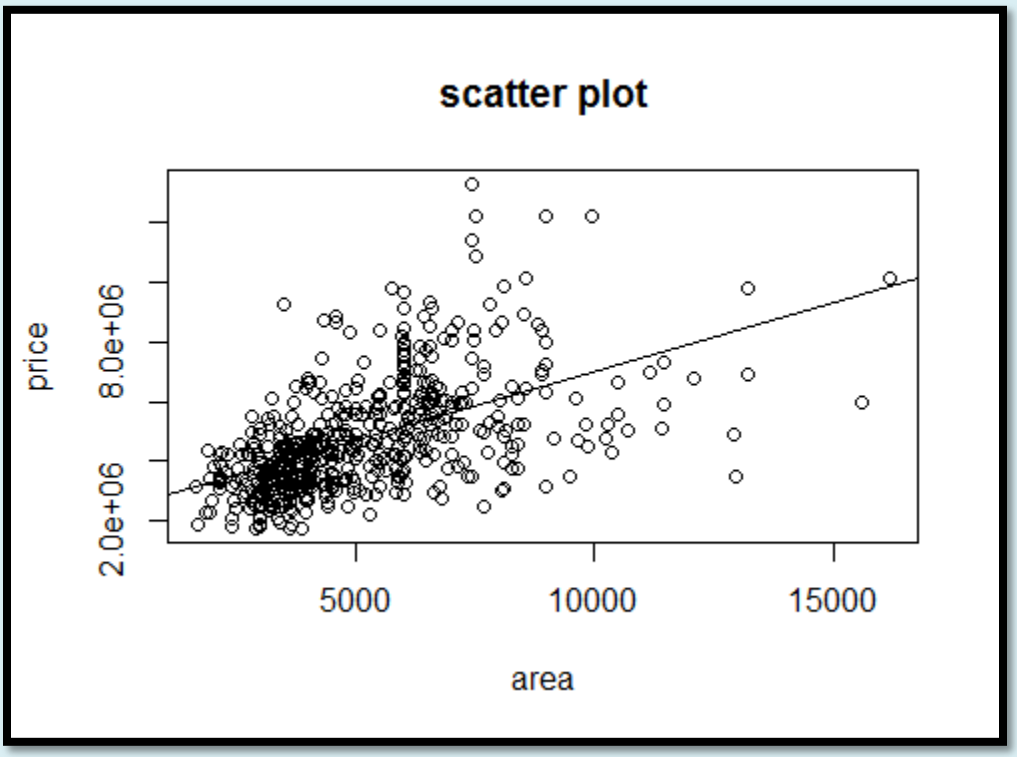



**> abline(lm(price~area))**

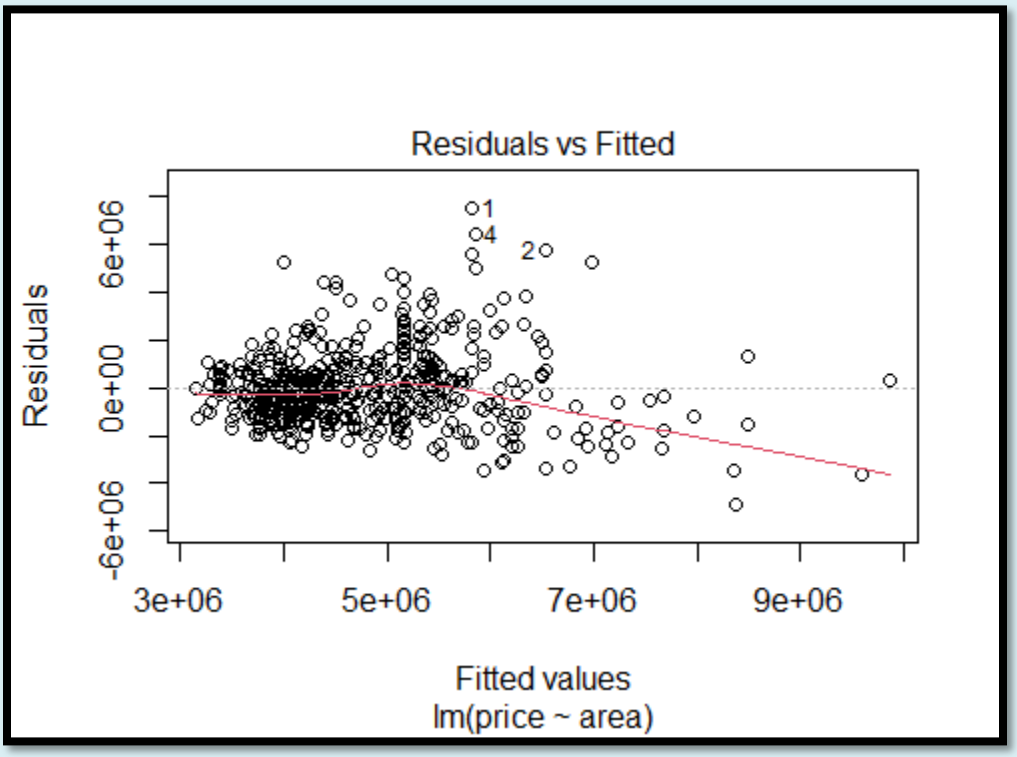


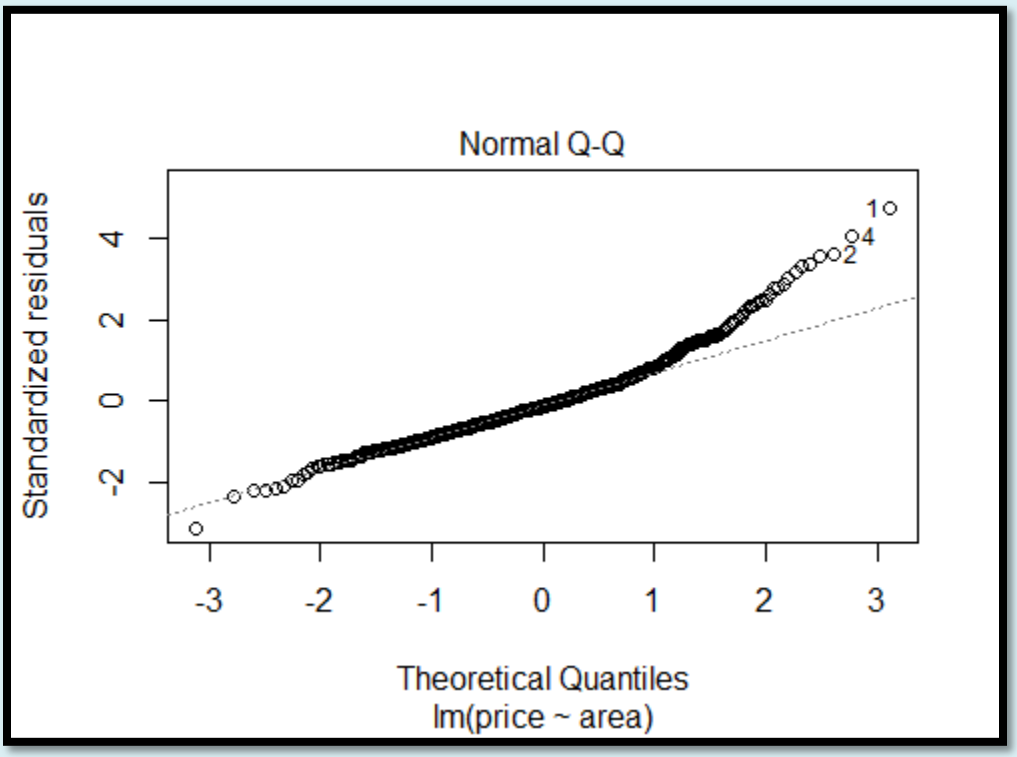
# Linear Regression using R- Programming

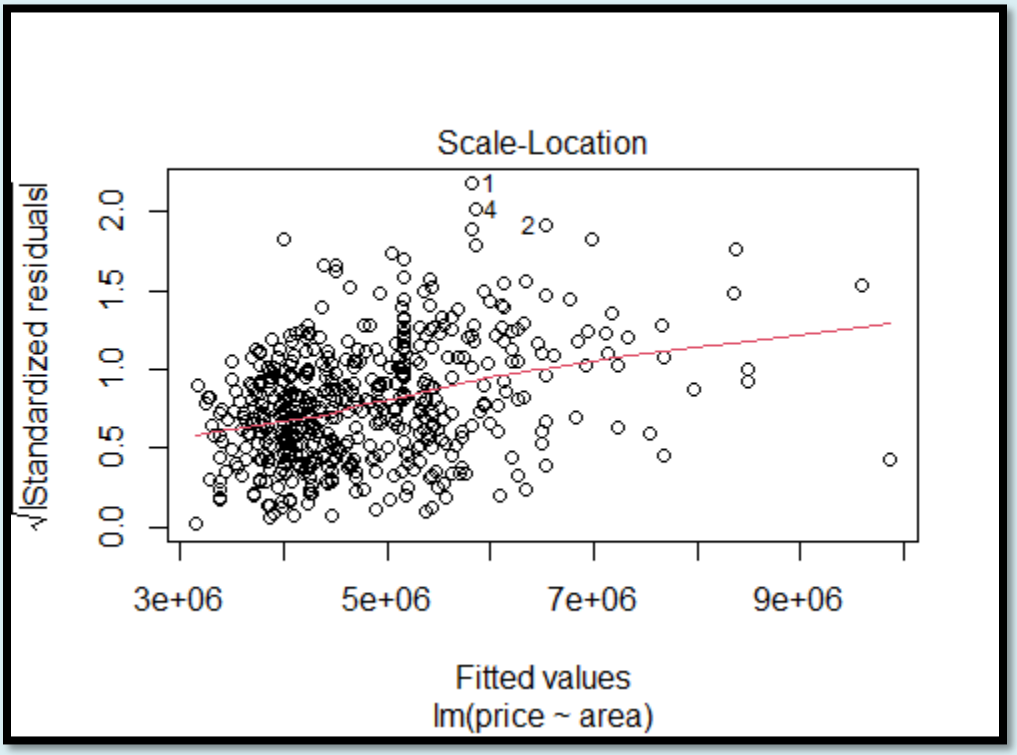
```
> #Simple Linear Regression
> #DV= Price; IDV= Area
> sreg= lm(price~area)
> summary(sreg)

Call:
lm(formula = price ~ area)

Residuals:
    Min      1Q  Median      3Q     Max
-4867112 -1022228  -200135   683027  7484838

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.387e+06  1.745e+05   13.68   <2e-16 ***
area        4.620e+02  3.123e+01   14.79   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1581000 on 543 degrees of freedom
Multiple R-squared:  0.2873,      Adjusted R-squared:  0.286
F-statistic: 218.9 on 1 and 543 DF,  p-value: < 2.2e-16

> plot(area,price, main = "scatter plot")
> abline(sreg)
```

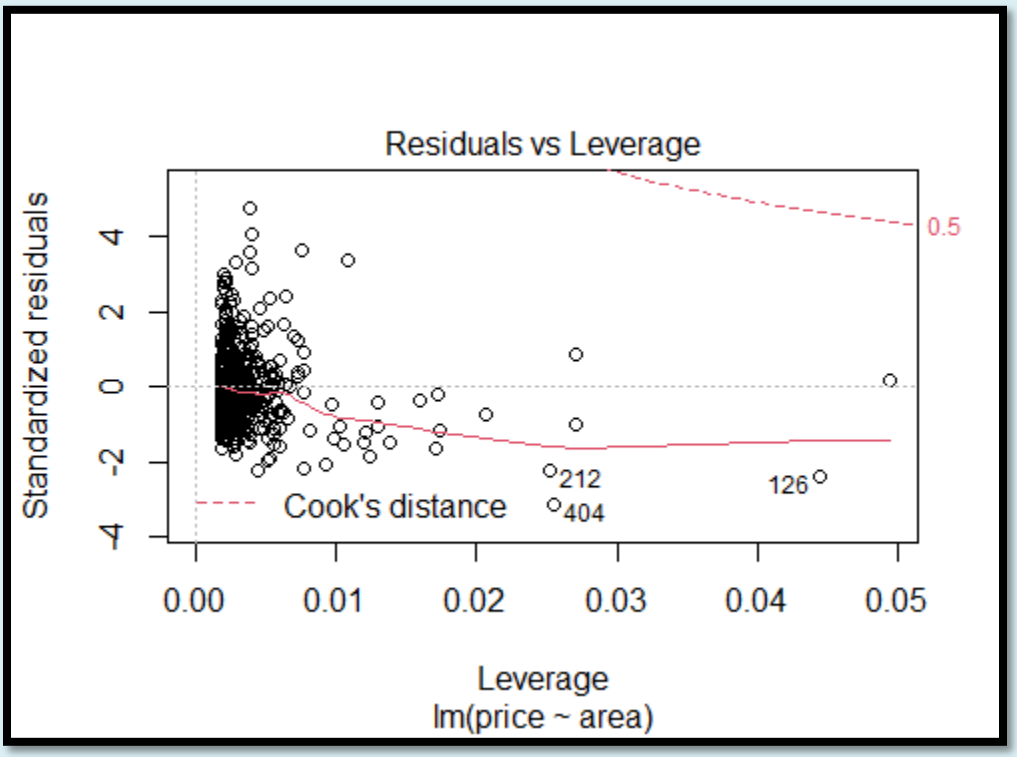

scatter plot
price
2.0e+06
8.0e+06
5000
10000
15000
area

**> plot(sreg)**


Residuals vs Fitted
Residuals
-6e+06
0e+00
6e+06
1
4
2
3e+06
5e+06
7e+06
9e+06
Fitted values
lm(price ~ area)

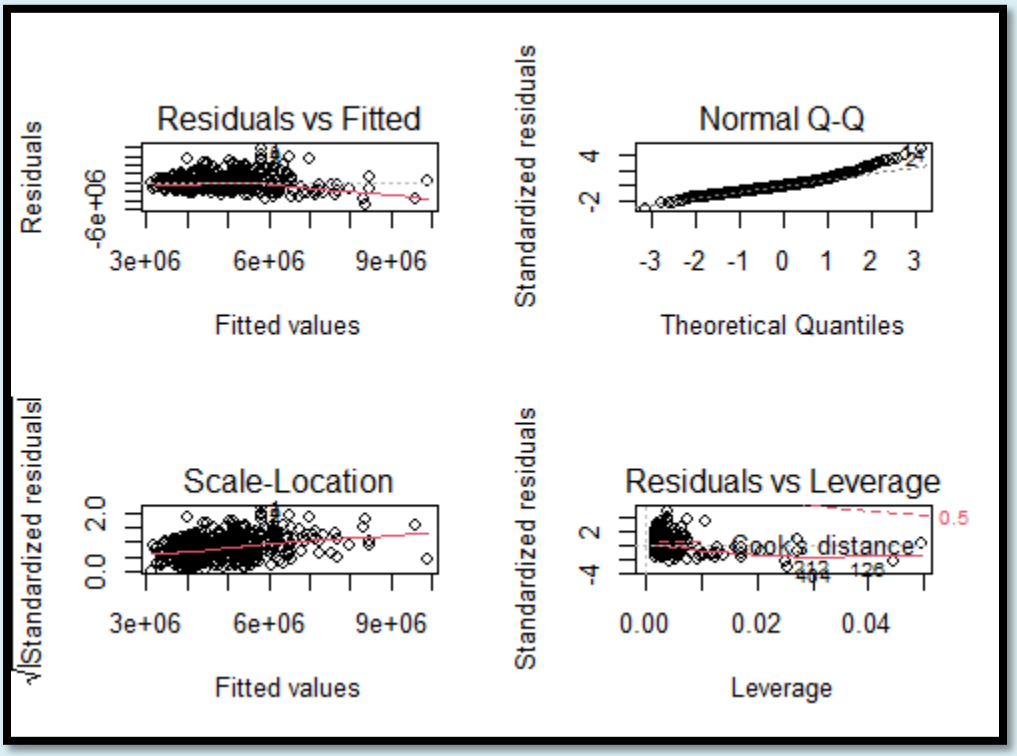

Normal Q-Q
Standardized residuals
-2
0
2
4
1
4
2
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(price ~ area)


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
1
4
2
3e+06
5e+06
7e+06
9e+06
Fitted values
lm(price ~ area)




**> par(mfrow=c(2,2))**

**> plot(sreg)**




**> #Multiple Linaer Regression**

**> mreg = lm(price~area+bedrooms+bathrooms+stories+mainroad+guestroom+basement+hotwaterheating+airconditioning)**

**> summary(mreg)**

Call:

lm(formula = price ~ area + bedrooms + bathrooms + stories + mainroad + guestroom + basement + hotwaterheating + airconditioning)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|
| -2790328 | -642601 | -20387 | 582368 | 5931692 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | -573514.10 | 247246.32 | -2.320 | 0.020738 | * |
| area | 303.30 | 24.47 | 12.392 | < 2e-16 | *** |
| bedrooms | 164594.14 | 76756.38 | 2.144 | 0.032453 | * |

bathrooms 1041793.18 109119.50 9.547 < 2e-16 ***

stories 445078.09 67889.83 6.556 1.30e-10 ***

mainroad 663026.18 147891.26 4.483 9.00e-06 ***

guestroom 318748.08 139464.36 2.286 0.022673 *

basement 507993.18 114628.11 4.432 1.13e-05 ***

hotwaterheating 919690.70 235479.04 3.906 0.000106 ***

airconditioning 951148.06 113990.62 8.344 6.14e-16 ***

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

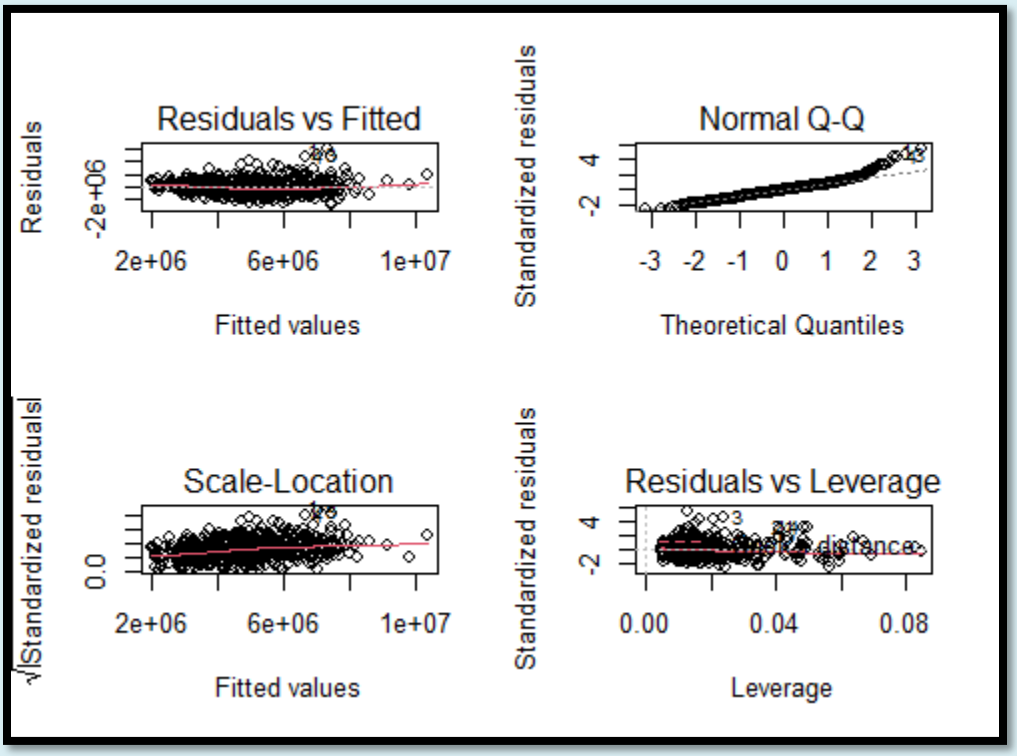
Residual standard error: 1134000 on 535 degrees of freedom

Multiple R-squared: 0.6385, Adjusted R-squared: 0.6324

F-statistic: 105 on 9 and 535 DF, p-value: < 2.2e-16

**> par(mfrow=c(2,2))**

**> plot(mreg)**



**> #Durbin Watson test for Autocorrelation**

**> install.packages("lmtest")**

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/Subho/Documents/R/win-library/4.0’

(as ‘lib’ is unspecified)

also installing the dependency ‘zoo’

trying URL 'http://cran.rstudio.com/bin/windows/contrib/4.0/zoo_1.8-8.zip'

Content type 'application/zip' length 1094462 bytes (1.0 MB)

downloaded 1.0 MB

trying URL 'http://cran.rstudio.com/bin/windows/contrib/4.0/lmtest_0.9-38.zip'

Content type 'application/zip' length 411871 bytes (402 KB)

downloaded 402 KB

package ‘zoo’ successfully unpacked and MD5 sums checked

package ‘lmtest’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Subho\AppData\Local\Temp\Rtmp4iikJ9\downloaded_packages

**> library(lmtest)**

Loading required package: zoo

Attaching package: ‘zoo’

The following objects are masked from ‘package:base’:

as.Date, as.Date.numeric

Warning messages:

1: package ‘lmtest’ was built under R version 4.0.3

2: package ‘zoo’ was built under R version 4.0.3

**> dwtest(mreg)**

Durbin-Watson test

data: mreg

DW = 1.1538, p-value < 2.2e-16

alternative hypothesis: true autocorrelation is greater than 0

**> #VIF to check multicolinearity**

**> install.packages("car")**

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/Subho/Documents/R/win-library/4.0’

(as 'lib' is unspecified)

trying URL 'http://cran.rstudio.com/bin/windows/contrib/4.0/car_3.0-10.zip'

Content type 'application/zip' length 1561500 bytes (1.5 MB)

downloaded 1.5 MB

package 'car' successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Subho\AppData\Local\Temp\Rtmp4iikJ9\downloaded_packages

**> library(car)**

Loading required package: carData

Warning message:

package 'car' was built under R version 4.0.3

**> car::vif(mreg)**

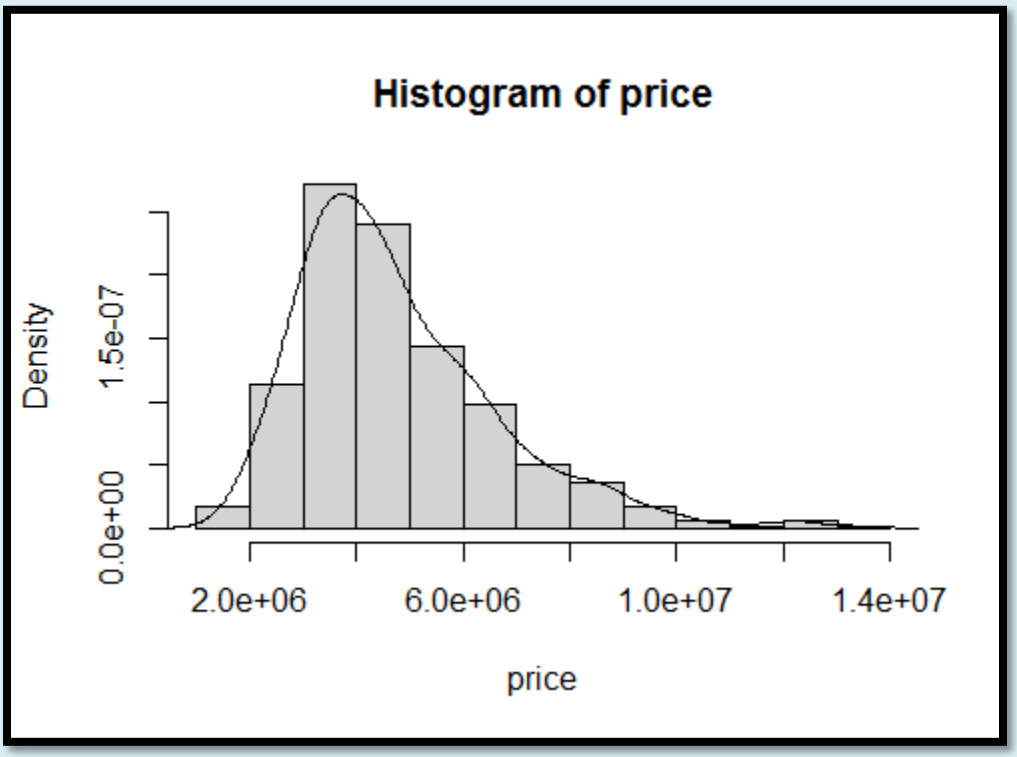
| area | bedrooms | bathrooms | stories | mainroad | guestroom |
|---|---|---|---|---|---|
| 1.193222 | 1.357446 | 1.271540 | 1.467056 | 1.124429 | 1.205831 |

| basement | hotwaterheating | airconditioning |
|---|---|---|
| 1.267445 | 1.028391 | 1.189285 |

**> hist(price)**

**> hist(price, freq = F)**

**> lines(density(price))**


Histogram of price
Density
0.0e+00
1.5e-07
2.0e+06
6.0e+06
1.0e+07
1.4e+07
price