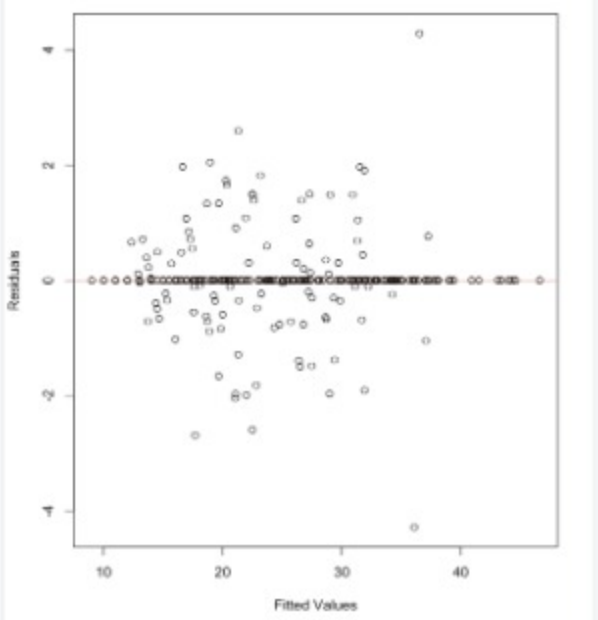
MTH-522_InClass

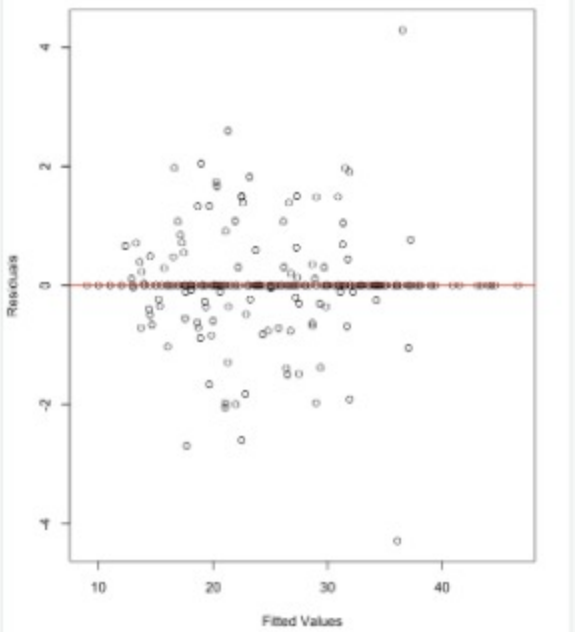
Banoth Jeevan Kumar

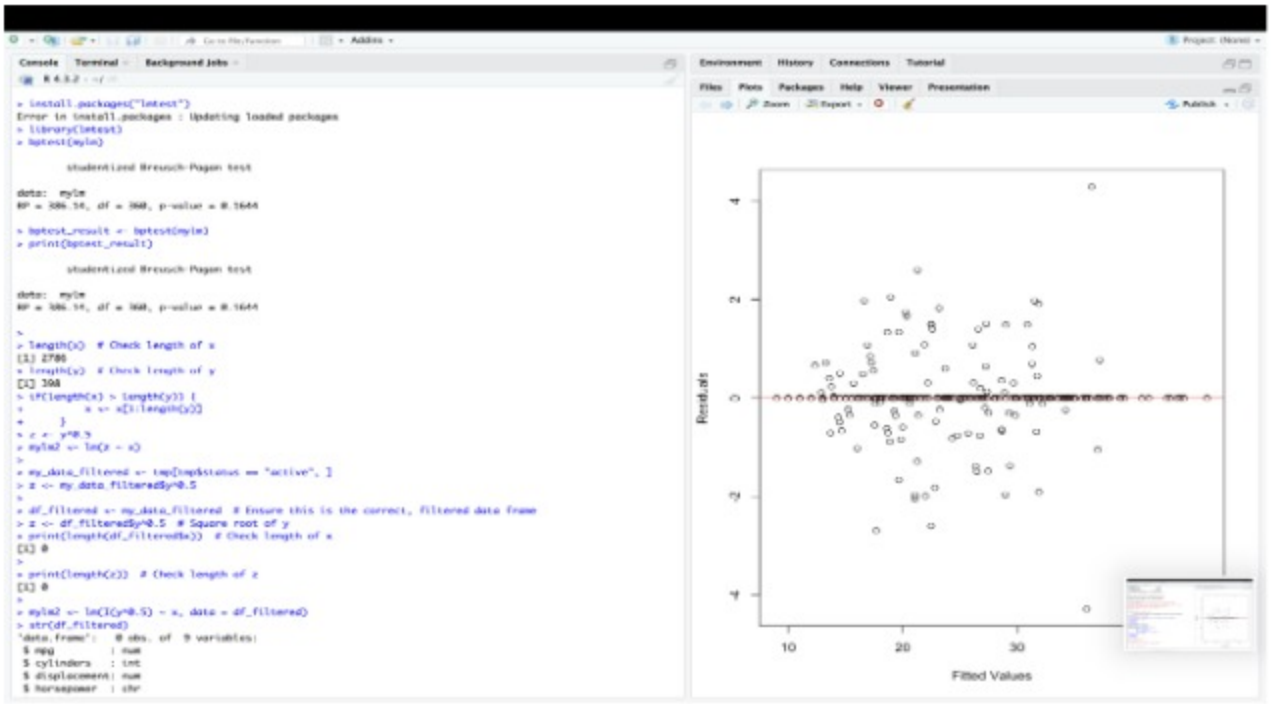
- The below are the pictures of In class activity that I performed in My laptop

```
R version 4.3.2 (2023-10-31) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> tmp <- read.csv("/Users/jeevankumarbanoth/Downloads/auto-mpg.csv", header=TRUE)
> x <- matrix(0, nrow(tmp), (p-1))
> for(i in 1:(p-1)) {
+           x[, i] <- tmp[, (i+1)]
+       }
> y <- tmp[,1]
> n <- nrow(tmp)
> p <- 8
> x <- matrix(0, nrow = n, ncol = p - 1)
> for(i in 1:(p - 1)) {
+           x[, i] <- tmp[, (i + 1)]
+       }
> mylm <- lm(tmp[, 1] ~ ., data = tmp[, -1])
> summary(mylm)

Call:
lm(formula = tmp[, 1] ~ ., data = tmp[, -1])

Residuals:
    Min      1Q  Median      3Q     Max
-4.283  0.000  0.000  0.000  4.283

Coefficients: (43 not defined because of singularities)
                                        Estimate Std. Error t value
(Intercept)                           24.322728  17.973000   1.353
cylinders                             -1.856873   1.087221  -1.708
displacement                          -0.019630   0.028203  -0.696
```

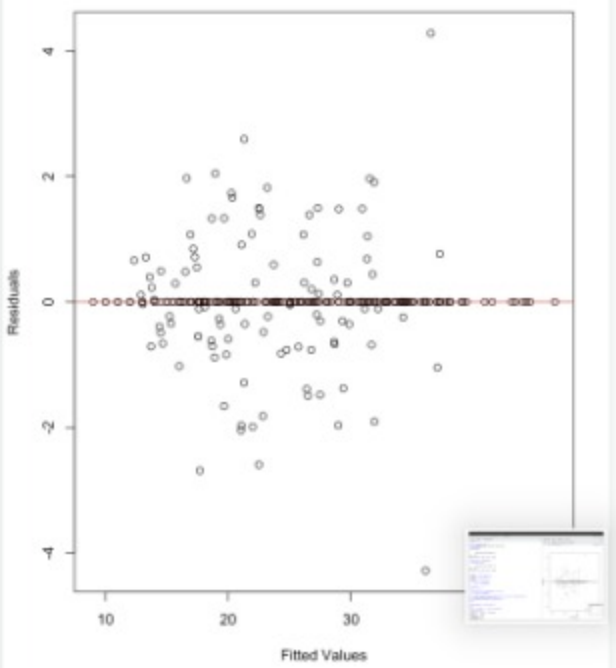



```
car.namedodge aspen                   0.76932
car.namedodge aspen 6                 0.96732
car.namedodge aspen se                0.84991
 [ reached getOption("max.print") -- omitted 204 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.873 on 37 degrees of freedom
Multiple R-squared:  0.9934,    Adjusted R-squared:  0.9796
F-statistic: 15.57 on 360 and 37 DF,  p-value: 6.884e-16

> install.packages("car")
Installing package into '/Users/jeevankumarbanoth/Library/R/arm64/4.3/library'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/macosx/big-sur-arm64/contrib/4.3/car_3.1-2.tgz'
Content type 'application/x-gzip' length 1710465 bytes (1.6 MB)
==================================================
downloaded 1.6 MB


The downloaded binary packages are in
        /var/folders/64/hzfllkgx5h15j_zjl8nwbssn0000gn/T//RtmpTqI9tF/downloaded_packages
> library(car)
Loading required package: carData
> plot(fitted(mylm), resid(mylm), xlab = "Fitted Values", ylab = "Residuals")
> abline(h = 0, col = "red")
> library(MASS)
> boxcox(mylm)
> ncvTest(mylm)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 25.8174, Df = 1, p = 3.7529e-07
> plot(fitted(mylm), resid(mylm),
+           xlab = "Fitted values",
+           ylab = "Residuals",
+           main = "Residuals vs Fitted Values")
> abline(h = 0, col = "red")
> library(MASS)
> boxcox(mylm)
> library(lmtest)
Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

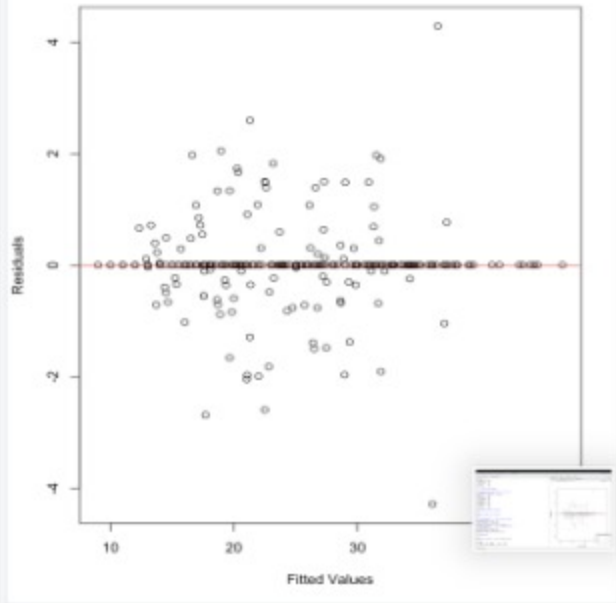


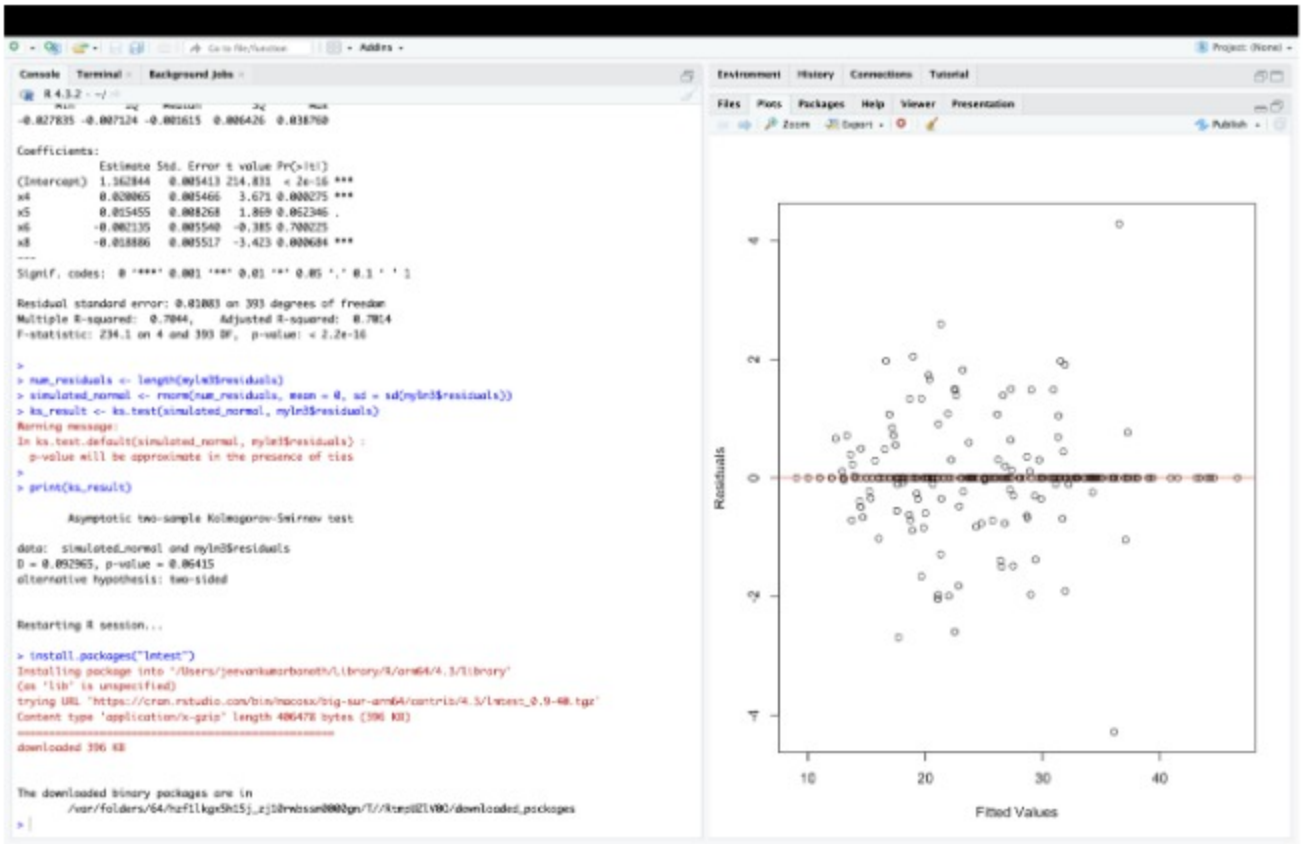

> install.packages("lmtest")
Error in install.packages : Updating loaded packages
> library(lmtest)
> bptest(mylm)
studentized Breusch-Pagan test
data: mylm
BP = 886.54, df = 368, p-value = 0.1644
> bptest_result <- bptest(mylm)
> print(bptest_result)
studentized Breusch-Pagan test
data: mylm
BP = 886.54, df = 368, p-value = 0.1644
> length(x) # Check length of x
[1] 2780
> length(y) # Check length of y
[1] 398
> if(length(x) > length(y)) {
+ x <- x[1:length(y)]
+ }
> z <- y^0.5
> mylm2 <- lm(y ~ x)
> my_data_filtered <- tmp[tmp$status == "active",]
> z <- my_data_filtered$y^0.5
> df_filtered <- my_data_filtered # Ensure this is the correct, filtered data frame
> z <- df_filtered$y^0.5 # Square root of y
> print(length(df_filtered$x)) # Check length of x
[1] 0
> print(length(z)) # Check length of z
[1] 0
> mylm2 <- lm(I(y^0.5) ~ x, data = df_filtered)
> str(df_filtered)
'data.frame': 0 obs. of 9 variables:
$ mpg : num
$ cylinders : int
$ displacement: num
$ horsepower : chr
Environment
History
Connections
Tutorial
Files
Plots
Packages
Help
Viewer
Presentation
Resid.dt
Fitted Values
10
20
30

```
 $ weight      : int
 $ acceleration: num
 $ model.year  : int
 $ origin      : int
 $ car.name    : chr
>
> z <- my_data_filtered$y^0.5
>
> my_data_filtered <- tmp[tmp$status == "active", ]
> str(my_data_filtered)
'data.frame':   0 obs. of  9 variables:
 $ mpg         : num
 $ cylinders   : int
 $ displacement: num
 $ horsepower  : chr
 $ weight      : int
 $ acceleration: num
 $ model.year  : int
 $ origin      : int
 $ car.name    : chr
>
> print(length(my_data_filtered$x))
[1] 0
>
> print(length(my_data_filtered$y))
[1] 0
>
> mylm2 <- lm(I(y^0.5) ~ x, data = my_data_filtered)
> ncvTest(mylm2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 20.59574, Df = 1, p = 5.6722e-06
> mylm3 <- lm(I(y^0.05) ~ x, data = my_data_filtered)
> ncvTest(mylm3)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 2.523721, Df = 1, p = 0.12741
>
> summary(mylm2)

Call:
lm(formula = I(y^0.5) ~ x, data = my_data_filtered)

Residuals:
     Min       1Q   Median       3Q      Max
-1.14403 -0.28765 -0.09418  0.27018  1.71134

Coefficients:
```




```
> summary(mylm2)

Call:
lm(formula = I(y^0.5) ~ x, data = my_data_filtered)

Residuals:
     Min       1Q   Median       3Q      Max
-1.14403 -0.28765 -0.09418  0.27018  1.71134

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.52665    0.22916  19.753  < 2e-16 ***
x4           0.86082    0.23140   3.717 0.000231 ***
x5           0.66622    0.35005   1.903 0.057742 .
x6          -0.07358    0.23456  -0.314 0.753925
x8          -0.67457    0.23357  -2.888 0.004090 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4583 on 393 degrees of freedom
Multiple R-squared:  0.6776,    Adjusted R-squared:  0.6743
F-statistic: 206.5 on 4 and 393 DF,  p-value: < 2.2e-16

> summary(mylm3)

Call:
lm(formula = I(y^0.05) ~ x, data = my_data_filtered)

Residuals:
      Min       1Q   Median       3Q      Max
-0.027835 -0.007124 -0.001615  0.006426  0.038760

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.162844   0.005413 214.831  < 2e-16 ***
x4           0.020065   0.005466   3.671 0.000275 ***
x5           0.015455   0.008268   1.869 0.062346 .
x6          -0.002135   0.005540  -0.385 0.700225
x8          -0.018886   0.005517  -3.423 0.000684 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01083 on 393 degrees of freedom
Multiple R-squared:  0.7044,    Adjusted R-squared:  0.7014
F-statistic: 234.1 on 4 and 393 DF,  p-value: < 2.2e-16

>
> num_residuals <- length(mylm3$residuals)
```





- **The below one is the markdown file of my work.**

MTH-522_InClass

Jeevan

2024-03-24

```
tmp <- read.csv("/Users/jeevankumarbanoth/Downloads/auto-mpg.csv",
header=TRUE)
```

```
x <- matrix(0, nrow(tmp), (p-1))
for(i in 1:(p-1)) {
     x[, i] <- tmp[, (i+1)]
 }
y <- tmp[,1]
n <- nrow(tmp)
p <- 8
```

```
x <- matrix(0, nrow = n, ncol = p - 1)
for(i in 1:(p - 1)) {
     x[, i] <- tmp[, (i + 1)]
 }
```

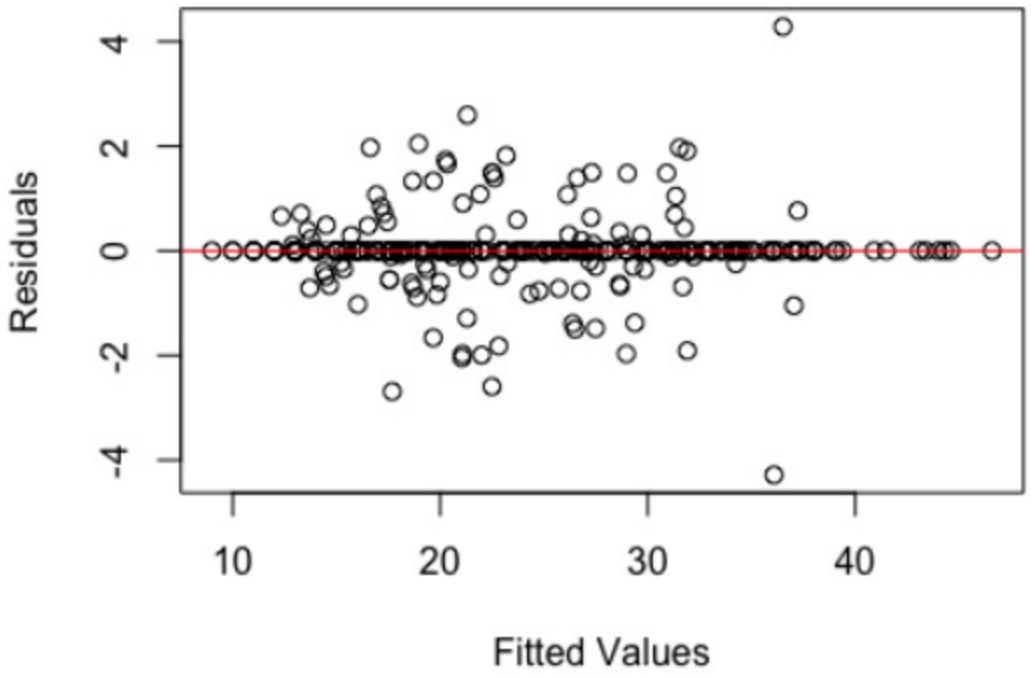
```
mylm <- lm(tmp[, 1] ~ ., data = tmp[, -1])
summary(mylm)
```

```
##
## Call:
## lm(formula = tmp[, 1] ~ ., data = tmp[, -1])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.283  0.000  0.000  0.000  4.283
##
## Coefficients: (43 not defined because of singularities)
##                                              Estimate Std. Error t value
## (Intercept)                                 24.322728  17.973000   1.353
## cylinders                                   -1.856873   1.087221  -1.708
## displacement                                -0.019630   0.028203  -0.696
## horsepower100                               -3.976684   3.190125  -1.247
## car.namevw rabbit custom                         NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.073 on 37 degrees of freedom
## Multiple R-squared:  0.9934, Adjusted R-squared:  0.9296
## F-statistic: 15.57 on 360 and 37 DF,  p-value: 6.084e-16
```

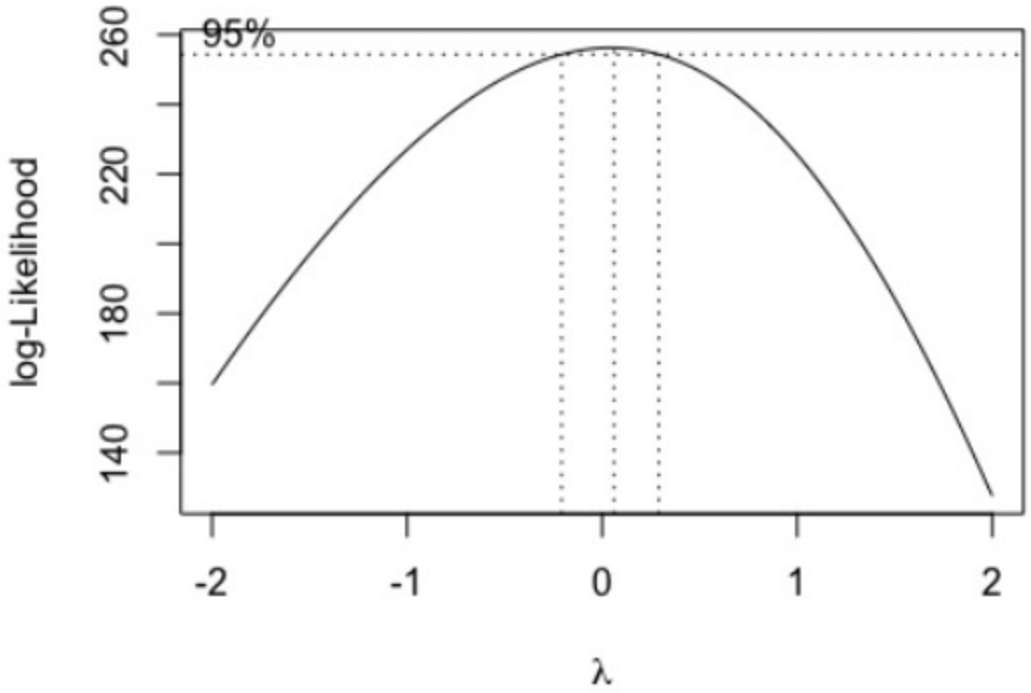
```
install.packages("car")
library(car)
```

```
## Loading required package: carData
```

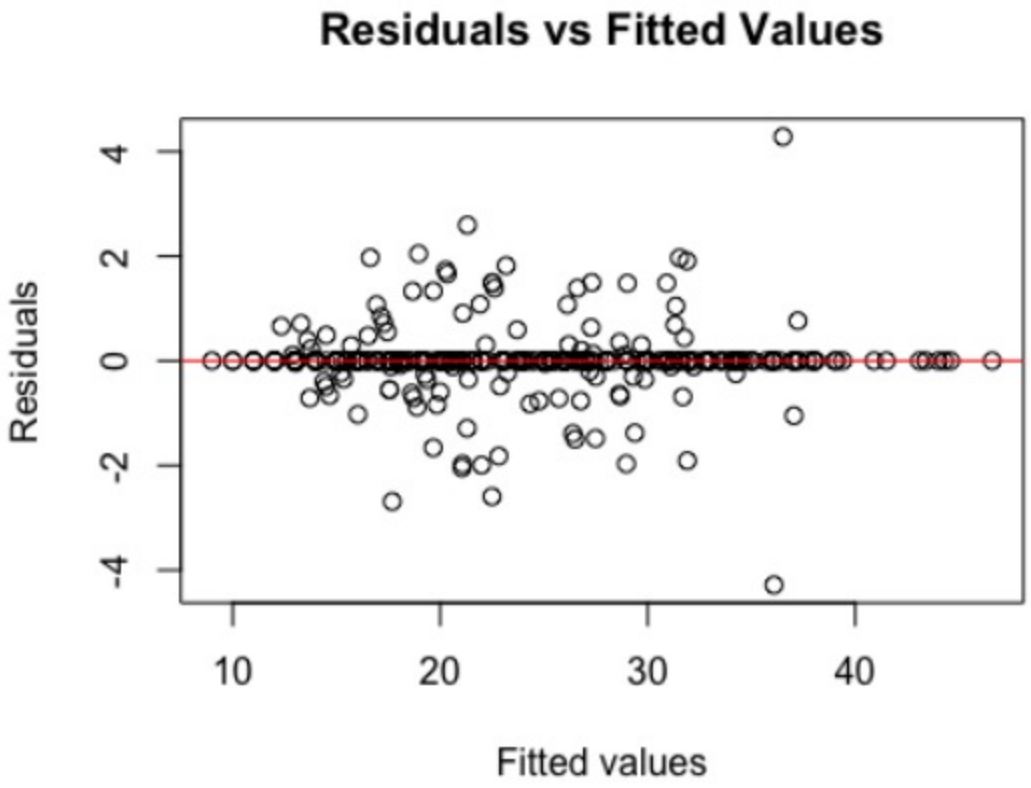
```
plot(fitted(mylm), resid(mylm), xlab = "Fitted Values", ylab = "Residuals")
abline(h = 0, col = "red")
```



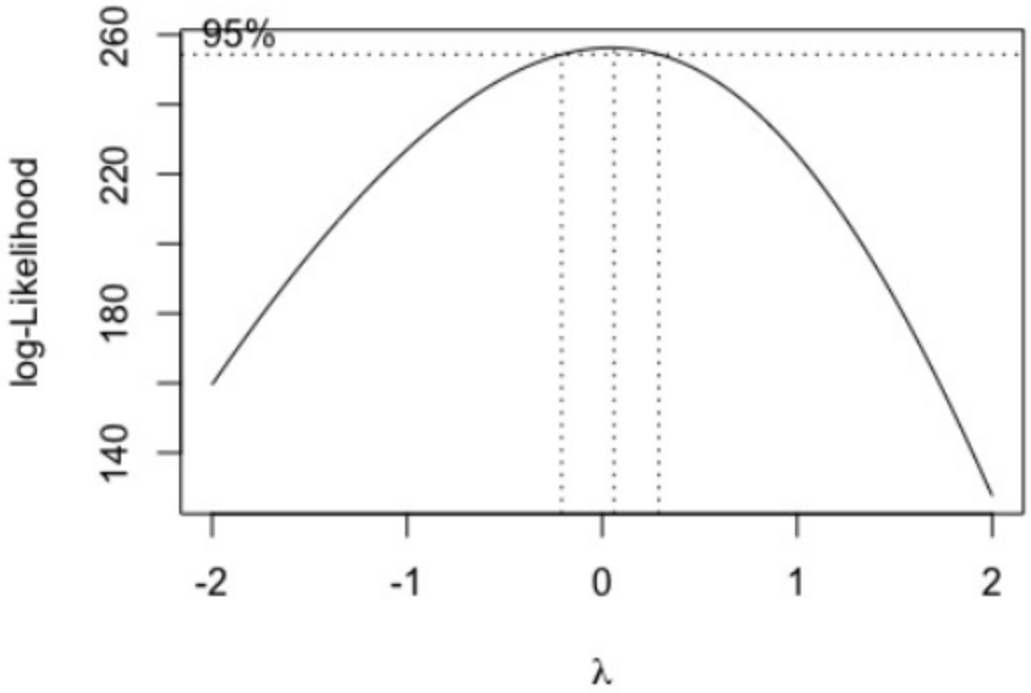

```
library(MASS)
boxcox(mylm)
```




```
ncvTest(mylm)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 25.8174, Df = 1, p = 3.7529e-07

plot(fitted(mylm), resid(mylm),
      xlab = "Fitted values",
      ylab = "Residuals",
      main = "Residuals vs Fitted Values")
abline(h = 0, col = "red")
```




```
library(MASS)
boxcox(mylm)
```




```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
bptest(mylm)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mylm
## BP = 386.14, df = 360, p-value = 0.1644
```

```
bptest_result <- bptest(mylm)
print(bptest_result)
```

```
##
##  studentized Breusch-Pagan test
##
```

```
## data:  mylm
## BP = 386.14, df = 360, p-value = 0.1644
```

```
length(x)  # Check length of x
length(y)  # Check length of y
if(length(x) > length(y)) {
+    x <- x[1:length(y)]
+ }
z <- y^0.5
mylm2 <- lm(z ~ x)
```

```
my_data_filtered <- tmp[tmp$status == "active", ]
z <- my_data_filtered$y^0.5

df_filtered <- my_data_filtered

z <- df_filtered$y^0.5  # Square root of y
print(length(df_filtered$x))  # Check length of x

## [1] 0

print(length(z))  # Check length of z

## [1] 0

str(df_filtered)

## 'data.frame':    0 obs. of  9 variables:
##  $ mpg         : num
##  $ cylinders   : int
##  $ displacement: num
##  $ horsepower  : chr
##  $ weight      : int
##  $ acceleration: num
##  $ model.year  : int
##  $ origin      : int
##  $ car.name    : chr

z <- my_data_filtered$y^0.5

my_data_filtered <- tmp[tmp$status == "active", ]
str(my_data_filtered)

## 'data.frame':    0 obs. of  9 variables:
##  $ mpg         : num
##  $ cylinders   : int
##  $ displacement: num
##  $ horsepower  : chr
##  $ weight      : int
##  $ acceleration: num
##  $ model.year  : int
##  $ origin      : int
##  $ car.name    : chr
```

```r
print(length(my_data_filtered$x))
```

```
## [1] 0
```

```r
print(length(my_data_filtered$y))
```

```
## [1] 0
```

```r
mylm2 <- lm(I(y^0.5) ~ x, data = df_filtered)
str(df_filtered)

z <- my_data_filtered$y^0.5

my_data_filtered <- tmp[tmp$status == "active", ]
str(my_data_filtered)

print(length(my_data_filtered$x))

print(length(my_data_filtered$y))

mylm2 <- lm(I(y^0.5) ~ x, data = my_data_filtered)
ncvTest(mylm2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 25.8174, Df = 1, p = 3.7529e-07
```

```r
mylm3 <- lm(I(y^0.05) ~ x, data = my_data_filtered)
ncvTest(mylm3)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 25.8174, Df = 1, p = 3.7529e-07
```

```r
summary(mylm2)
```

```
Call:
lm(formula = I(y^0.5) ~ x, data = my_data_filtered)

Residuals:
   Min     1Q  Median     3Q    Max
-1.14403 -0.28765 -0.09418  0.27018  1.71134

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.52665    0.22916  19.753  < 2e-16 ***
x4           0.86002    0.23140   3.717 0.000231 ***
x5           0.66622    0.35005   1.903 0.057742 .
x6          -0.07358    0.23456  -0.314 0.753925
x8          -0.67457    0.23357  -2.888 0.004090 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4583 on 393 degrees of freedom
```

```
Multiple R-squared:  0.6776,    Adjusted R-squared:  0.6743
F-statistic: 206.5 on 4 and 393 DF,  p-value: < 2.2e-16
```

```
summary(mylm)
```

```
Call:
lm(formula = I(y^0.05) ~ x, data = my_data_filtered)

Residuals:
     Min       1Q   Median       3Q      Max
-0.027835 -0.007124 -0.001615  0.006426  0.038760

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.162844   0.005413 214.831  < 2e-16 ***
x4           0.020065   0.005466   3.671 0.000275 ***
x5           0.015455   0.008268   1.869 0.062346 .
x6          -0.002135   0.005540  -0.385 0.700225
x8          -0.018886   0.005517  -3.423 0.000684 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01083 on 393 degrees of freedom
Multiple R-squared:  0.7044,    Adjusted R-squared:  0.7014
F-statistic: 234.1 on 4 and 393 DF,  p-value: < 2.2e-16
```

```
num_residuals <- length(mylm$residuals)
simulated_normal <- rnorm(num_residuals, mean = 0, sd = sd(mylm$residuals))
ks_result <- ks.test(simulated_normal, mylm$residuals)

## Warning in ks.test.default(simulated_normal, mylm$residuals): p-value will
be
## approximate in the presence of ties

print(ks_result)

##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  simulated_normal and mylm$residuals
## D = 0.40201, p-value < 2.2e-16
## alternative hypothesis: two-sided
```