LAB7 – NON-LINEAR REGRESSION

A.Mayuri(2348133)
[COMPANY NAME] [Company address]

Lab7 - Non linear Regression

A.Mayuri(2348133)

2024-01-05

Problem

By using the inbuilt data set “pressure” in R,

1)fit a suitable linear regression model that relates pressure and temperature.

2)perform residual analysis and comment about the possibility of modelling non linear regression model for the data set.

Introduction:

Non-linear regression model:

Nonlinear regression is a mathematical function that uses a generated line – typically a curve – to fit an equation to some data.

The sum of squares is used to determine the fitness of a regression model, which is computed by calculating the difference between the mean and every point of data.

Analysis:

Importing Data set

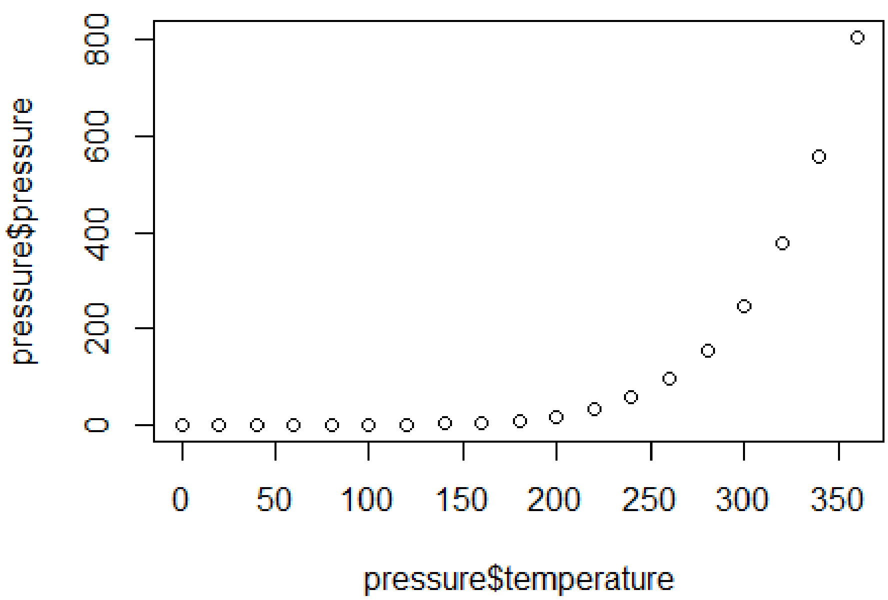
```
data("pressure")
pressure
```

```
##    temperature pressure
## 1            0   0.0002
## 2           20   0.0012
## 3           40   0.0060
## 4           60   0.0300
## 5           80   0.0900
## 6          100   0.2700
## 7          120   0.7500
## 8          140   1.8500
## 9          160   4.2000
## 10         180   8.8000
## 11         200  17.3000
## 12         220  32.1000
## 13         240  57.0000
## 14         260  96.0000
## 15         280 157.0000
```

```
## 16        300 247.0000
## 17        320 376.0000
## 18        340 558.0000
## 19        360 806.0000
```

Plotting and correlation

```
plot(pressure$temperature,pressure$pressure)
```




```
cor(pressure$temperature,pressure$pressure)

## [1] 0.7577923
```

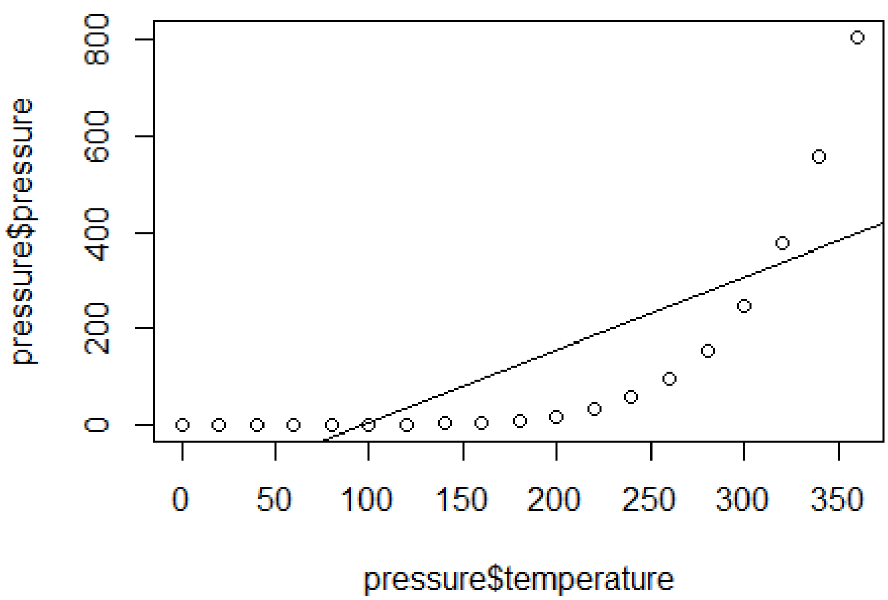
model

```
coke=lm(pressure$pressure~pressure$temperature)
summary(coke)

##
## Call:
## lm(formula = pressure$pressure ~ pressure$temperature)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -158.08 -117.06  -32.84   72.30  409.43
##
```

```
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -147.8989    66.5529  -2.222 0.040124 *
## pressure$temperature    1.5124     0.3158   4.788 0.000171 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.8 on 17 degrees of freedom
## Multiple R-squared:  0.5742, Adjusted R-squared:  0.5492
## F-statistic: 22.93 on 1 and 17 DF,  p-value: 0.000171
```

```
plot(pressure$temperature,pressure$pressure)
abline(coke)
```



<u>Interpretation:</u> we observe that the model is not a very good fit, thus let us try fitting a polynomial linear model. Also we can only fit a model of maximum polynomial degree 2 since we observe only 1 bend (minima) in the model.

Polynomial form of linear model

```
# using indicator method
hub=lm(pressure$pressure~pressure$temperature+I(pressure$temperature^2))
summary(hub)
```

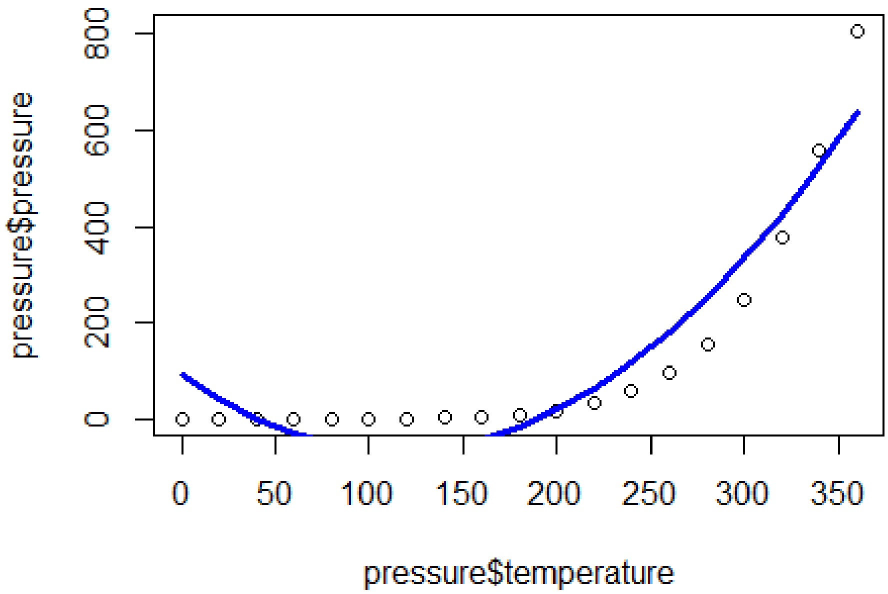
```
##
## Call:
## lm(formula = pressure$pressure ~ pressure$temperature + I(pressure$temperature^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -95.142 -54.391  -1.353  48.238 170.374
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               91.154379  46.262513   1.970 0.066354 .
## pressure$temperature      -2.706167   0.595775  -4.542 0.000333 ***
## I(pressure$temperature^2)  0.011718   0.001597   7.336 1.67e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74.42 on 16 degrees of freedom
## Multiple R-squared:  0.9024, Adjusted R-squared:  0.8902
## F-statistic:    74 on 2 and 16 DF,  p-value: 8.209e-09
```

Interpretation: we observe that all regression coefficients are significant. thus let us proceed to checking the validity of the model by performing the residual analysis.

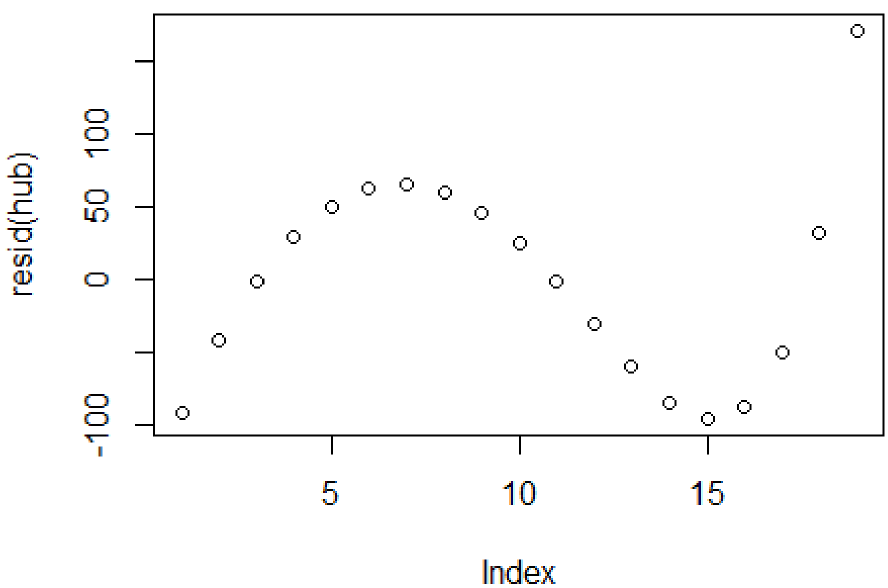
Residual Analysis.

```
plot(pressure$temperature,pressure$pressure)

lines(smooth.spline(pressure$temperature,predict(hub)),col="blue",lwd=3)# fitting the new model to the scatter plot
```




```
plot(resid(hub))# residual plot
```



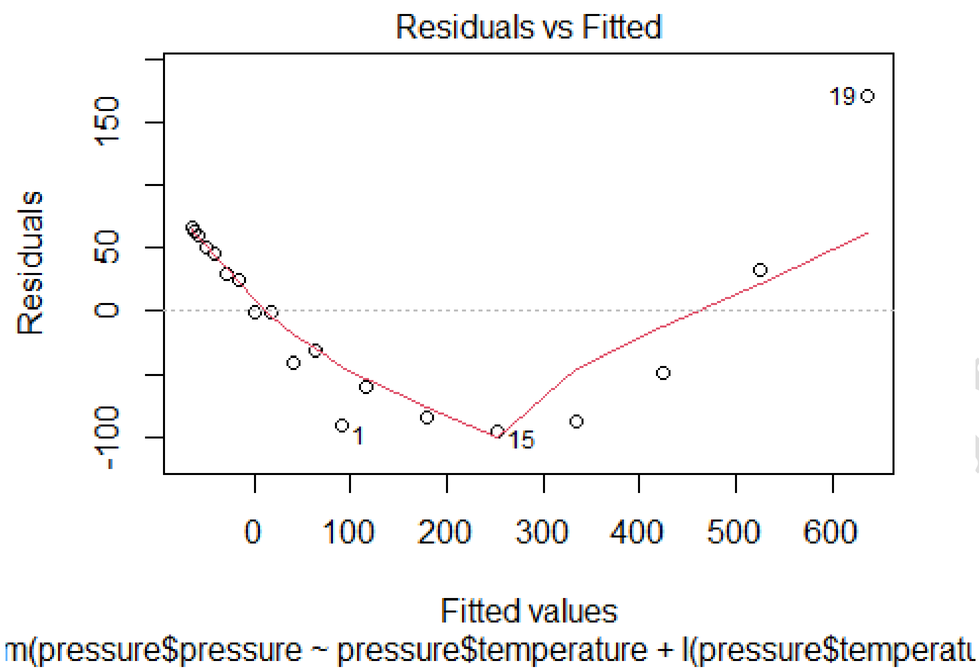
Interpretation: the model is a better fit, but the residuals are not randomn. (defies assumption of const variance) .

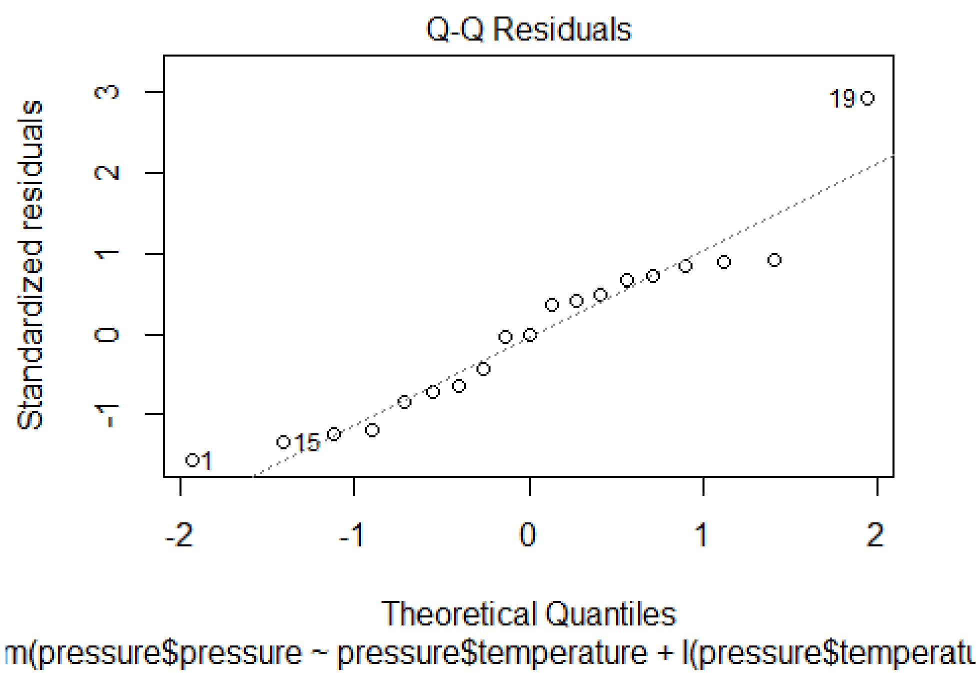
Normality Test

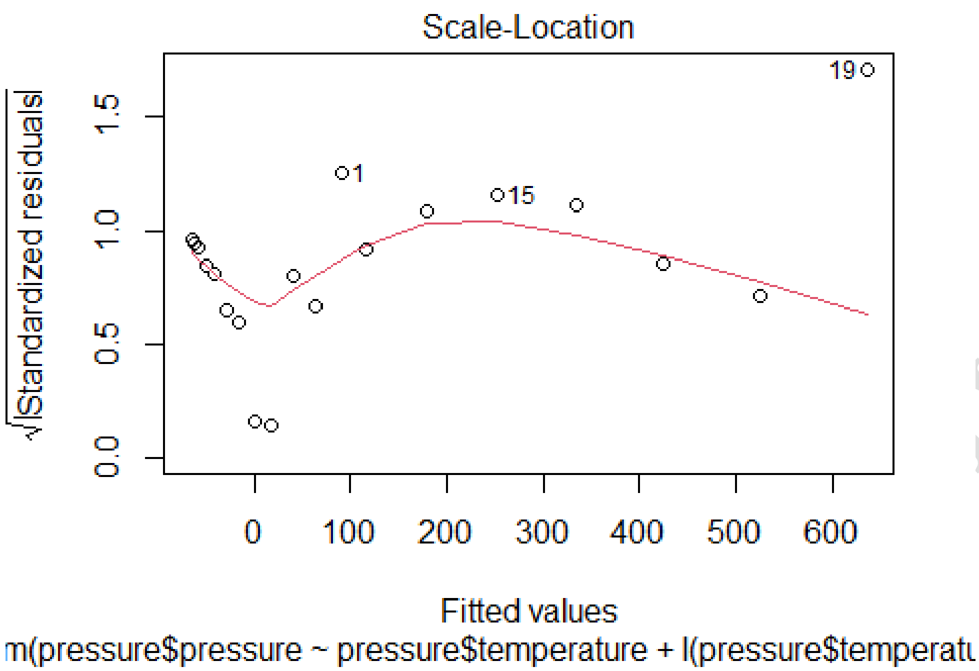
```
# to check for res vales (Normality check)
re1=rstandard(hub)
re1
```

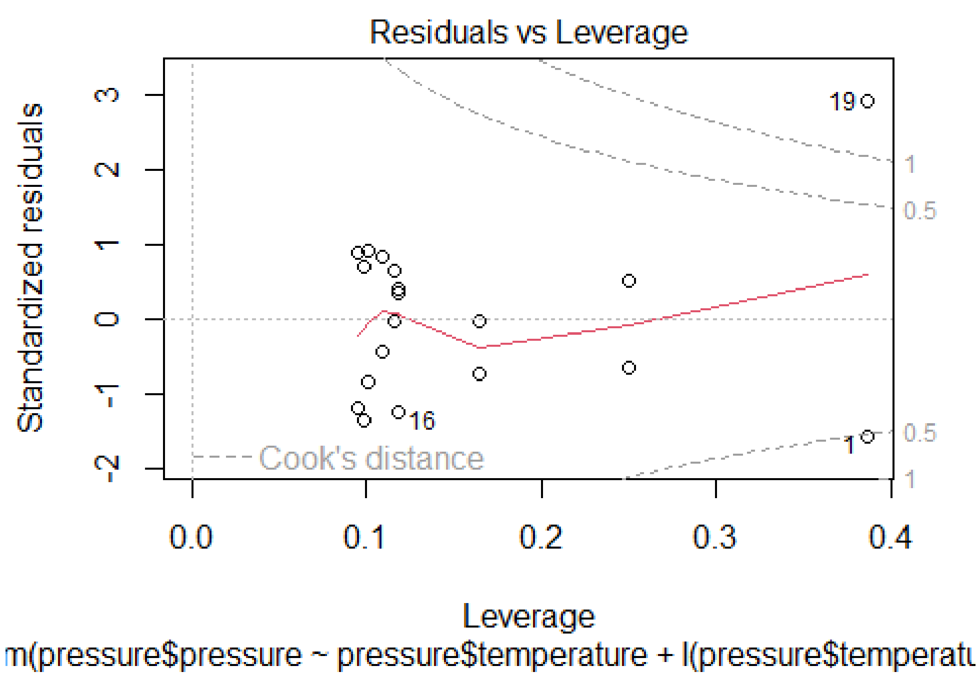
```
##           1           2           3           4           5           6 
## -1.56380948 -0.64736155 -0.02428165  0.41590439  0.71368868  0.88361058 
##           7           8           9          10          11          12 
##  0.92956879  0.85268591  0.65821226  0.35909478 -0.01933957 -0.43948670 
##          13          14          15          16          17          18 
## -0.84533112 -1.18250841 -1.34640718 -1.24444570 -0.72264345  0.50134024 
##          19 
##  2.92288522
```

```
plot(hub)
```


Residuals vs Fitted
Residuals
150
50
0
-100
19
1
15
0
100
200
300
400
500
600
Fitted values
m(pressure$pressure ~ pressure$temperature + I(pressure$temperatu


Q-Q Residuals
Standardized residuals
3
2
1
0
-1
19
15
1
-2
-1
0
1
2
Theoretical Quantiles
m(pressure$pressure ~ pressure$temperature + I(pressure$temperatu






```
shapiro.test(re1)

##
##  Shapiro-Wilk normality test
```

```
##
## data:  re1
## W = 0.92626, p-value = 0.1477
```

Interpretation: Shapiro wilk test : at 0.05 level of confidence we observe that p value > alpha thus we accept null hypothesis and conclude that it follows normality.

Hypothesis testing for constant variance :

test for errors have constant variance through errors. use the fitted model.

H0: error have const variance

H1: error have not const variance

```
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

bptest(hub)

##
##  studentized Breusch-Pagan test
##
## data:  hub
## BP = 7.4533, df = 2, p-value = 0.02407
```

Interpretation: we observe p value (0.024)<alpha(0.05) thus we reject null hypothesis and state that the model does not have a constant variance.

the model is a better fit, but the residuals are not randomn. (defies assumption of const variance) thus we cannot do linear regression. and step into non-linear regression model.

Conclusion:

Thus we need to shift from linear regression to non-linear regression for the data-set.

Nonlinear regression is a procedure that will determine the values for the parameters that minimizes the sum of the squares of the distances of the data points to the curve for the equation of interest. In other words, given the data point y and the corresponding point on the curve ˆy, nonlinear regression will minimize the sum of the their differences squared: SS=sum[(y − yˆ)^2]

Note that the term (y − yˆ) is simply the residual for each data point. This type of method is referred to as a least-squares method and is only applicable if the uncertainty is normally distributed

Computer Algorithm:

Computer algorithms are used to solve nonlinear regression problems. They work by iterating through values of the constants in the designated model until the residual sum of squares is minimized. Because of this, the programs have to be given an initial guess of the parameters from which to start their iterations. The initial guess is often based on rough calculations or intuition. For most cases, the algorithm should be able to determine the parameters regardless of the initial guess. However, it is possible for the algorithm to go in the wrong direction and never converge on a solution or for it to converge on the wrong solution. In short, be careful with your choice for the initial guess and always check your result to make sure it behaves reasonably.

References:

https://sites.duke.edu/bossbackup/files/2013/02/NonLinSummary.pdf

https://corporatefinanceinstitute.com/resources/data-science/nonlinear-regression/