Statistical Learning Laboratory 1

Ankit Rathi
Anastas Liti

The dataset that we have used to solve the given task is part 12.

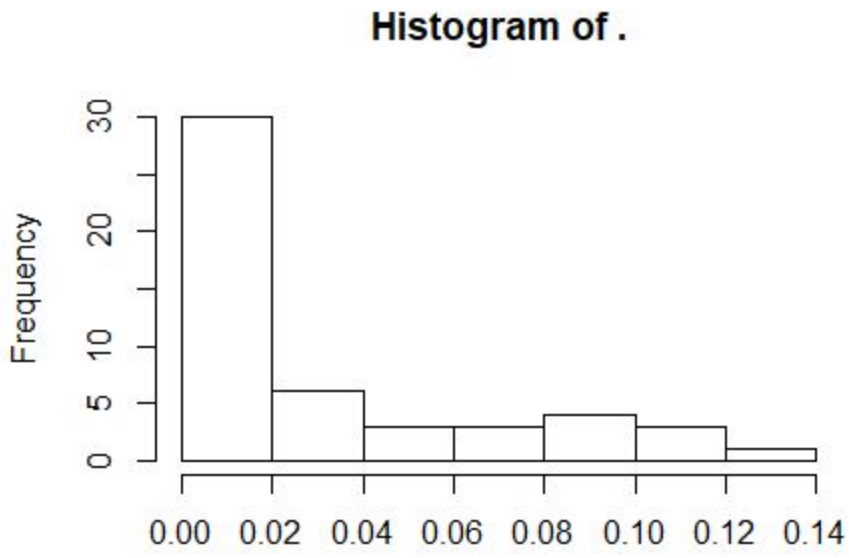
Part 2 :

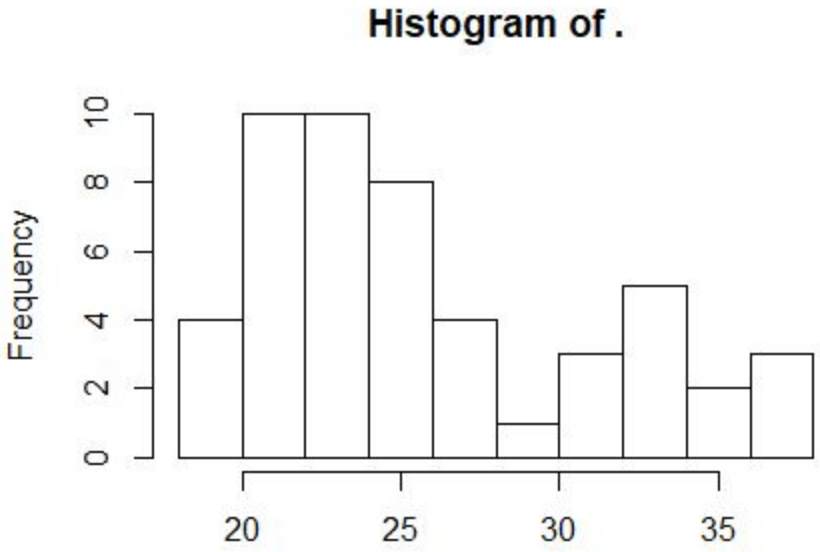
1. Shapiro-wilk Normality test
   data: df$density
   W = 0.75457, p-value = 9.117e-08

Comment-- it is not a normal dist as p<0.005

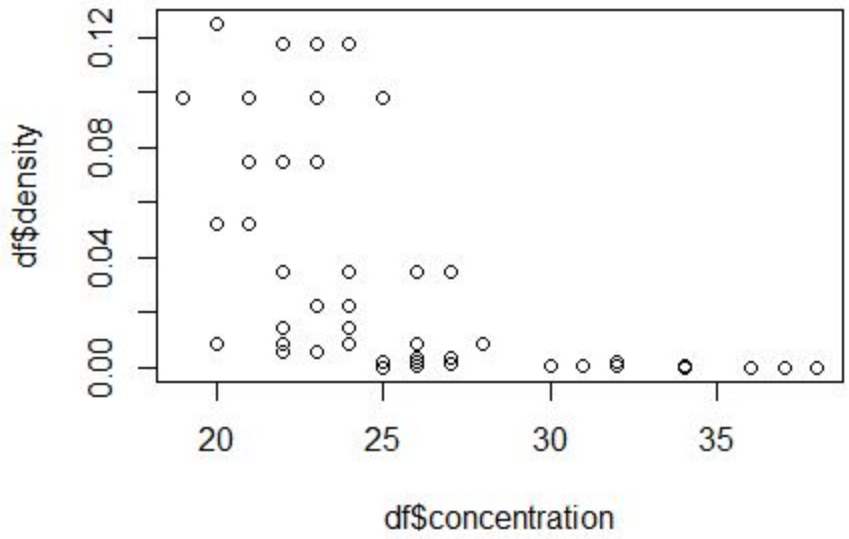
data: df$concentration
W = 0.89393, p-value = 0.0003046

Comment--it is not a normal dist as p<0.005





2. Scatterplot



3. because the output of p value is 7.737e-06 <0.05 we can say that correlation exist correlation value is -0.5861335.
   because the output of p value is 1.45e-12<0.05 we can say that correlation exist correlation value is 0.8071393.

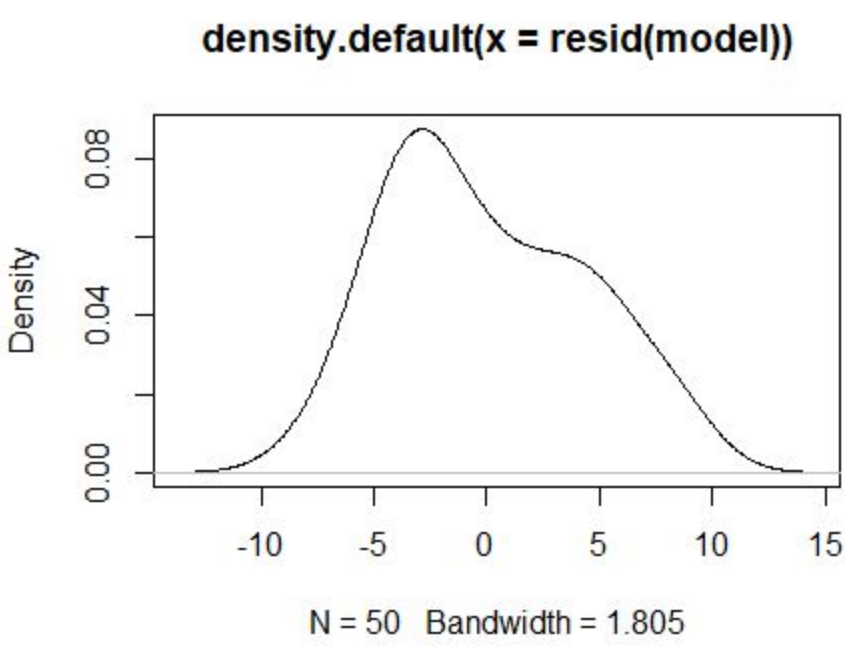
4. comment on scatterplot and correlation
   as it can be seen from the scatterplot and the correlation values the variables are negatively correlated to each other and can be written as a negative linear relationship between the two variables.

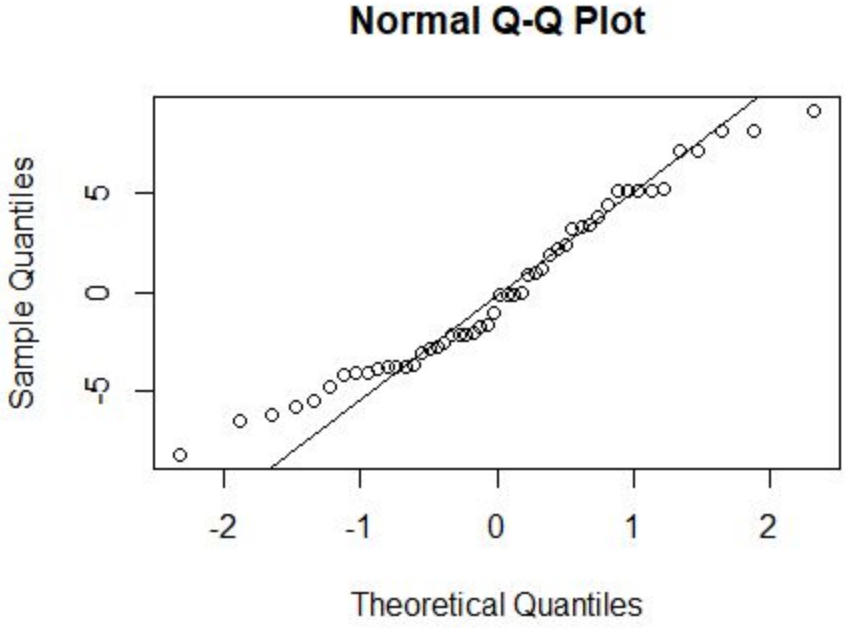
5. Coefficient of linear data

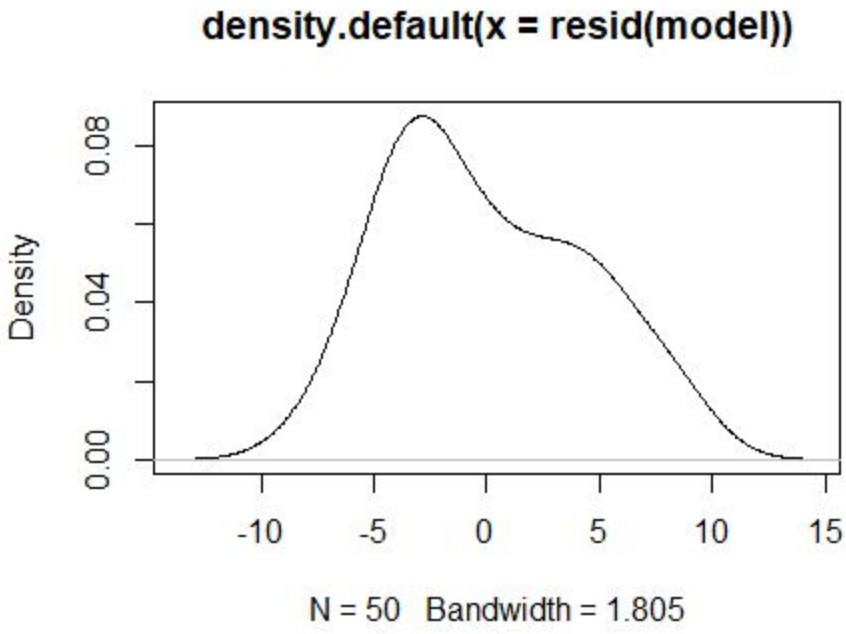
   Coefficients:
   (Intercept) df$density
   28.87 -78.47
   coeff =-78.47

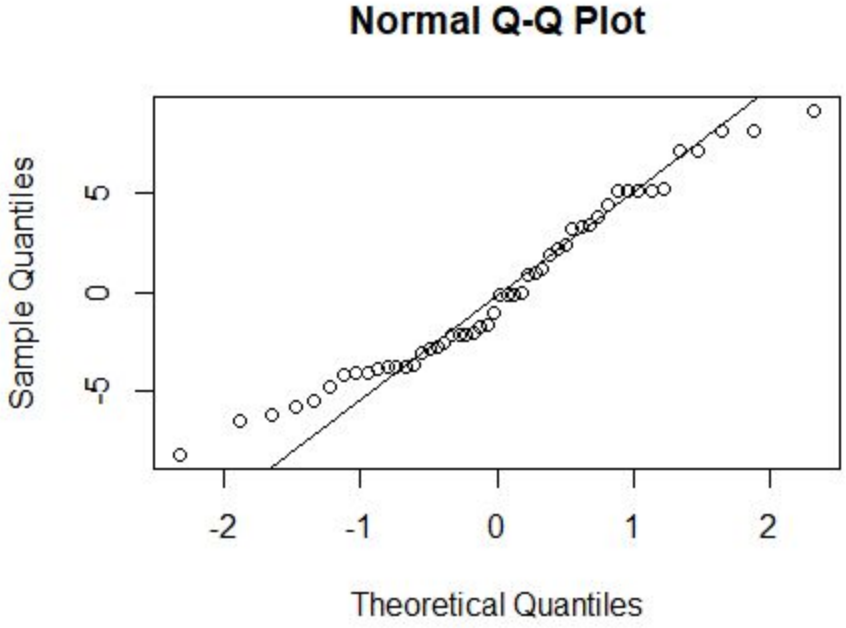
   Residual Plot

```
> resid(model) #List of residuals
         1          2          3          4          5          6          7          8          9
-2.1238092 -1.0226672  5.1614262 -0.1603812 -4.1018370 -1.7775256 -2.7152218  5.1473804  1.8396188
        10         11         12         13         14         15         16         17         18
-3.7549314 -6.1813241 -3.1018370 -0.0226672 -4.7549314  4.3458945 -6.4479591 -4.1238092  7.1337269
        19         20         21         22         23         24         25         26         27
 8.1337269 -2.1813241  3.2847782  1.1473804  7.1337269 -5.4479591 -0.1238092 -3.8713735 -8.1813241
        28         29         30         31         32         33         34         35         36
 2.1844958  0.9342491 -3.7549314 -1.6134485 -2.8155042  3.3458945 -0.1813241  3.1614262 -4.1813241
        37         38         39         40         41         42         43         44         45
 5.1473804  3.8396188  5.1258016  0.8761908  5.1275279 -5.7579097 -3.7579097 -2.0226672  9.1258016
        46         47         48         49         50
 8.1268217  2.3458945 -3.7152218 -2.6134485 -2.1603812
```


density.default(x = resid(model))
Density
0.00
0.04
0.08
-10
-5
0
5
10
15
N = 50 Bandwidth = 1.805

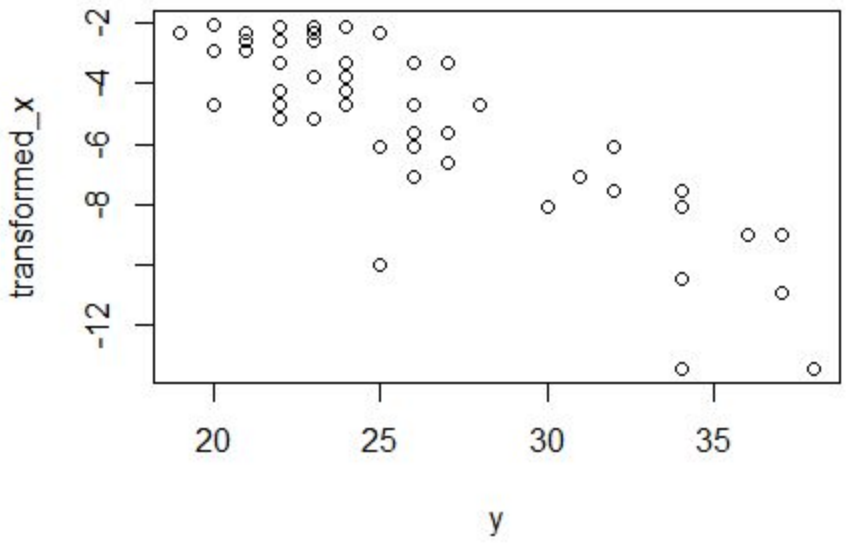

Normal Q-Q Plot
Sample Quantiles
-5
0
5
-2
-1
0
1
2
Theoretical Quantiles


density.default(x = resid(model))
Density
0.00
0.04
0.08
-10
-5
0
5
10
15
N = 50 Bandwidth = 1.805


Normal Q-Q Plot
Sample Quantiles
-5
0
5
-2
-1
0
1
2
Theoretical Quantiles

6.

```
> y<-df$concentration
> x<-df$density
> x
 [1] 0.035052 0.074573 0.000455 0.098306 0.022588 0.001233 0.002027 0.000276 0.098306 0.052497 0.008831
[12] 0.022588 0.074573 0.052497 0.117502 0.005433 0.035052 0.000102 0.000102 0.008831 0.002027 0.000276
[23] 0.000102 0.005433 0.035052 0.000037 0.008831 0.000749 0.125000 0.052497 0.003324 0.000749 0.117502
[34] 0.008831 0.000455 0.008831 0.000276 0.098306 0.000001 0.035052 0.000023 0.014227 0.014227 0.074573
[45] 0.000001 0.000014 0.117502 0.002027 0.003324 0.098306
> y
 [1] 24 22 34 21 23 27 26 34 23 21 22 24 23 20 24 22 22 36 37 26 32 30 36 23 26 25 20 31 20 21 27 26 23
[34] 28 32 24 34 25 34 27 34 22 24 21 38 37 22 25 26 19
>
```



7. Pearson correlation coefficient

Because the output of p value is 6.695e-15 <0.05 we can say that correlation exist correlation value is -0.849047.

```
> pearson_corr

        Pearson's product-moment correlation

data:  y and transformed_x
t = -11.134, df = 48, p-value = 6.695e-15
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9118880 -0.7473117
sample estimates:
      cor
-0.849047
```

Spearman rank correlation coefficient

Because the output of p value is 1.45e-12 <0.05 we can say that correlation exist correlation value is -0.8071393.

```
> spearman_corr

        Spearman's rank correlation rho

data:  y and transformed_x
S = 37634, p-value = 1.45e-12
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.8071393
```

8.

a.

```
> cor.diff.test(x,transformed_x,y,y,method="pearson")  #no sign difference
cor1: r=0.726, p=2.49e-09, n=50
cor2: r=1, p=0, n=50
diffence: p(one-sided)=0, p(two-sided)=0
$cor1
$cor1$estimate
[1] 0.7256493

$cor1$p.value
[1] 2.494812e-09

$cor1$n
[1] 50


$cor2
$cor2$estimate
[1] 1

$cor2$p.value
[1] 0

$cor2$n
[1] 50


$p.value.twosided
[1] 0

$p.value.onesided
[1] 0
```

B.

```
> cor.diff.test(x,transformed_x,y,y,method = 'spearman')
cor1: r=1, p=0, n=50
cor2: r=1, p=0, n=50
diffence: p(one-sided)=0, p(two-sided)=0
$cor1
$cor1$estimate
[1] 1

$cor1$p.value
[1] 0

$cor1$n
[1] 50


$cor2
$cor2$estimate
[1] 1

$cor2$p.value
[1] 0

$cor2$n
[1] 50


$p.value.twosided
[1] 0

$p.value.onesided
[1] 0
```

9.

Linear model after transformation

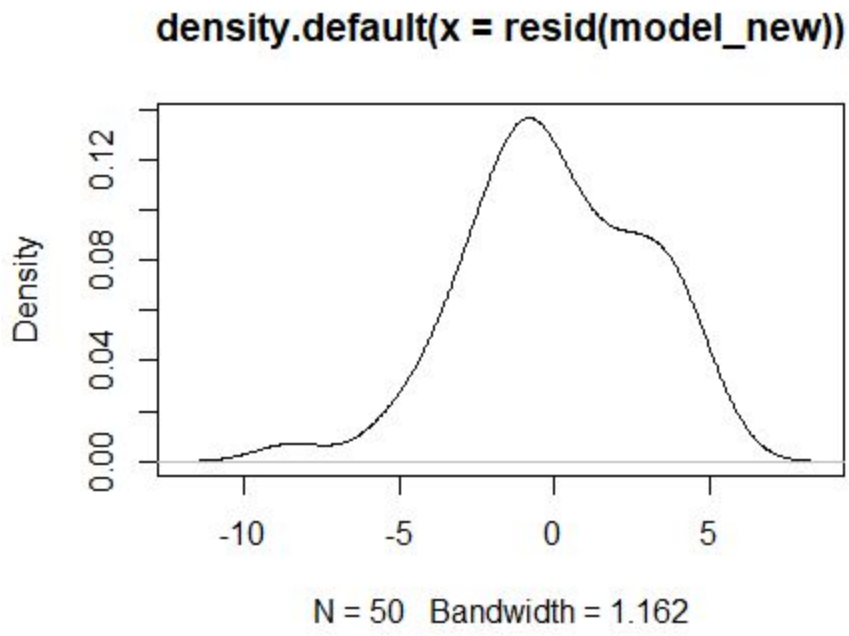
```
Call:
lm(formula = y ~ transformed_x)

Coefficients:
  (Intercept)  transformed_x
       18.068         -1.542
```

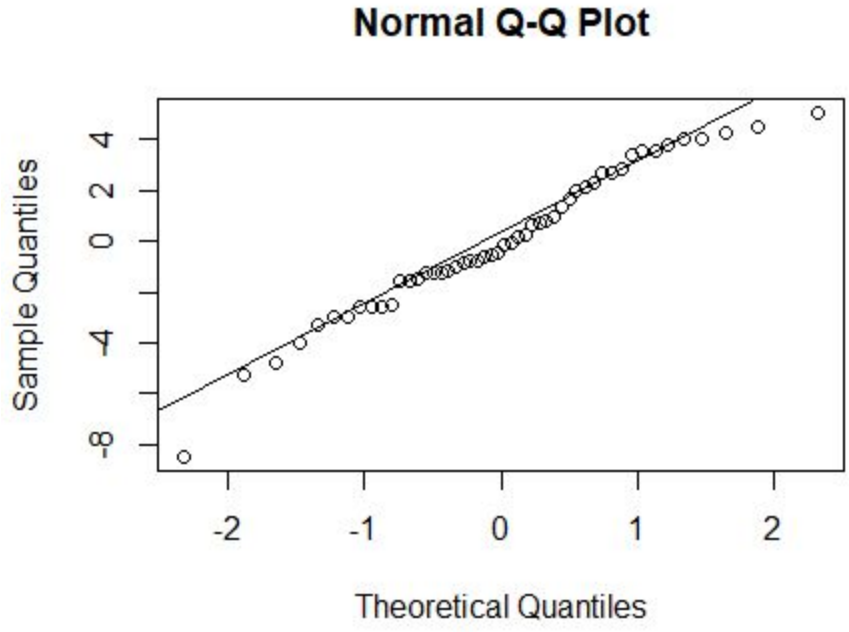
```
Call:
lm(formula = y ~ transformed_x)

Residuals:
    Min      1Q  Median      3Q     Max
-8.4661 -1.5631 -0.3267  2.2199  5.0346

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    18.0676     0.8574   21.07  < 2e-16 ***
transformed_x  -1.5417     0.1385  -11.13  6.7e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.89 on 48 degrees of freedom
Multiple R-squared:  0.7209,    Adjusted R-squared:  0.7151
F-statistic:   124 on 1 and 48 DF,  p-value: 6.695e-15
```

```
Call:
lm(formula = df$concentration ~ df$density)

Residuals:
   Min     1Q Median     3Q    Max
-8.181 -3.745 -0.602  3.331  9.126

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.8743     0.7905  36.526  < 2e-16 ***
df$density  -78.4682    15.6558  -5.012 7.74e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.432 on 48 degrees of freedom
Multiple R-squared:  0.3436,    Adjusted R-squared:  0.3299
F-statistic: 25.12 on 1 and 48 DF,  p-value: 7.737e-06
```

Comment: both p value and f value suggest the model have become better.

11.





A quantile normal plot - good for checking normality

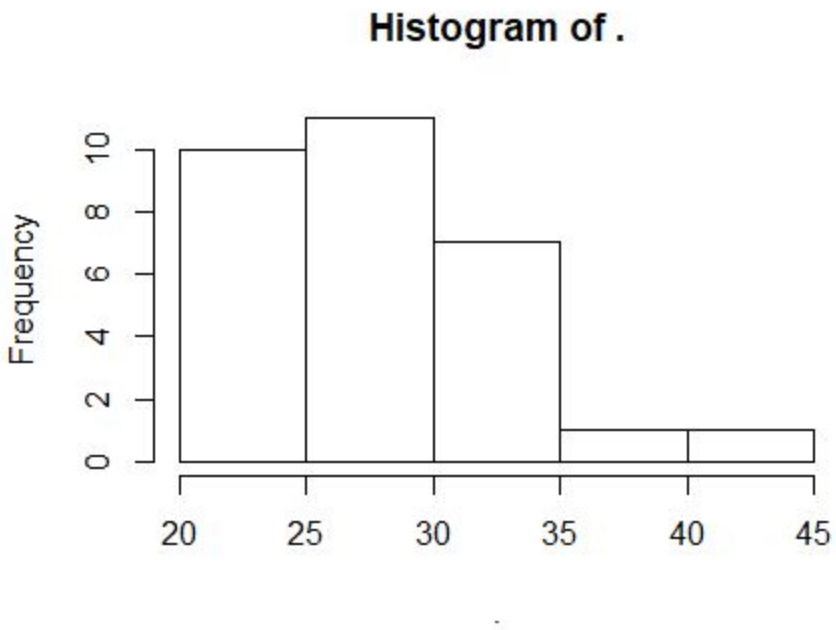
12.
coefficient of determination ---- 0.7208808
coefficient of determination of old model --- 0.3435525

13.
model performance has become better after the transfromation as can be seen form Q12.

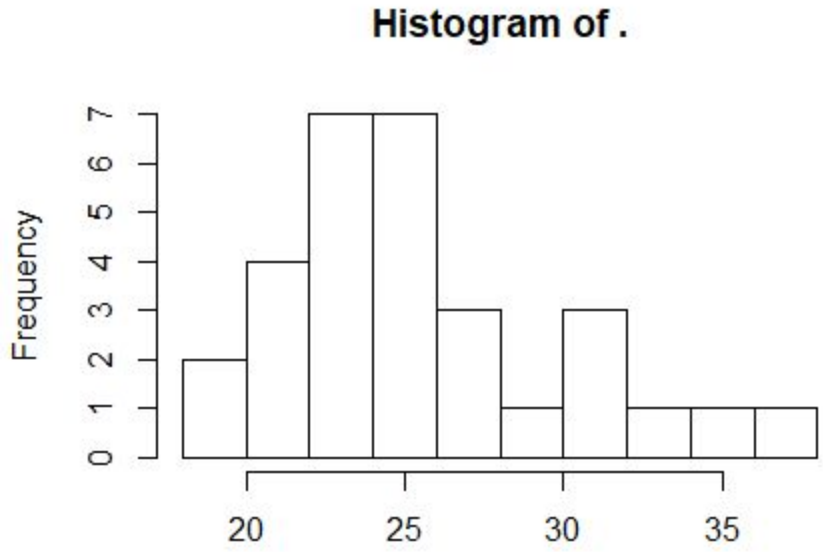
Part 1 :

1. Shapiro-Wilk Test

```
> shapiro.test(df$dose)

        Shapiro-Wilk normality test

data:  df$dose
W = 0.93352, p-value = 0.06095

> shapiro.test(df$concentration)

        Shapiro-Wilk normality test

data:  df$concentration
W = 0.94573, p-value = 0.1298
```

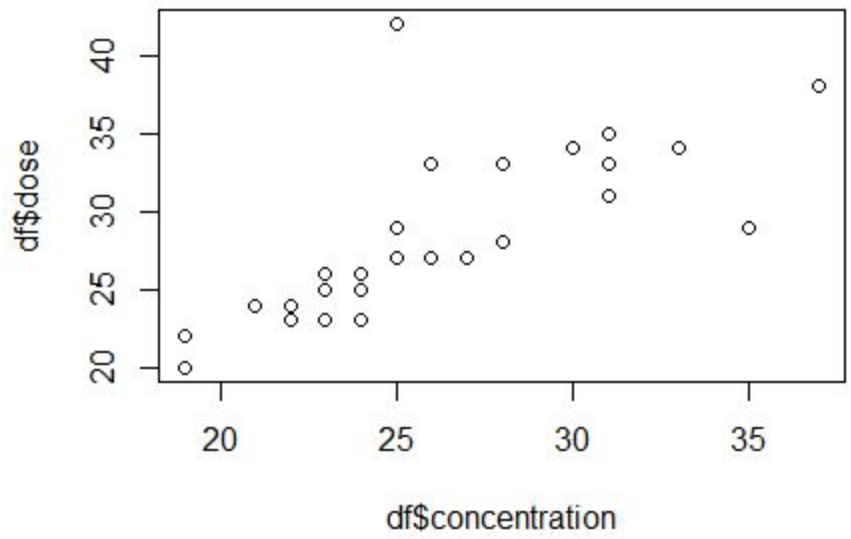


DOSE



CONCENTRATION

2. Scatterplot



3.

```
> pearson_corr

        Pearson's product-moment correlation

data:  df$dose and df$concentration
t = 5.6646, df = 28, p-value = 4.53e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5028893 0.8636558
sample estimates:
      cor
0.7307623
```

because the output of p value is 4.53e-0.6<0.05 we can say that correlation exist.
Cor value -- 0.7307623

```
> spearman_corr

        Spearman's rank correlation rho

data:  df$dose and df$concentration
S = 553.11, p-value = 2.074e-10
alternative hypothesis: true rho is not equal to 0
sample estimates:
     rho
0.876949
```

output of p value is 2.074e-10<0.05 we can say that correlation exist.
Cor value -- 0.876949

4.

as it can be seen from the scatterplot and the correlation values the variables are negatively correlated to each other and can be written as a negative linear relationship between the two variables.

5.

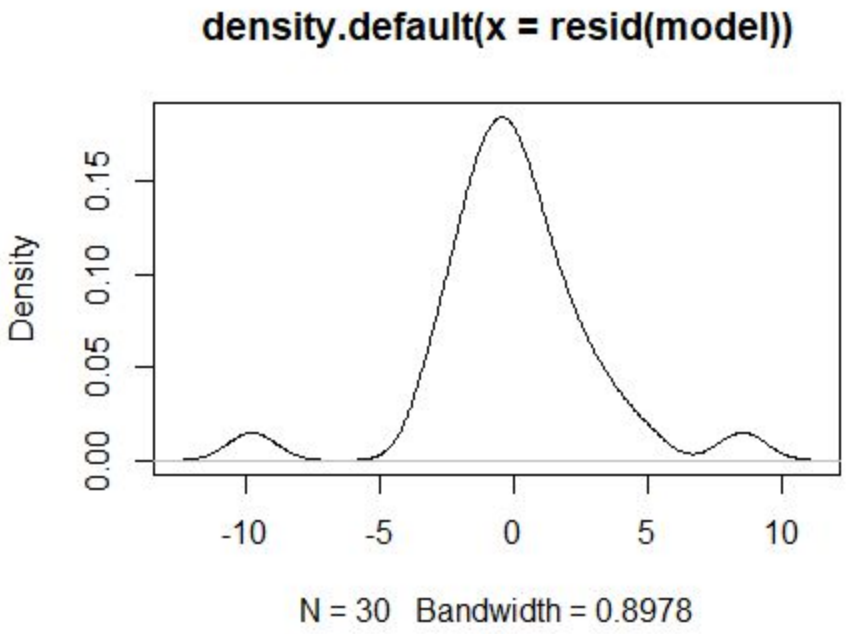
6.

7.

```
> summary(model_new)$r.squared
[1] 0.7208808
> mmary(model)$r.squared
Error in mmary(model) : could not
> summary(model)$r.squared
[1] 0.5340135
```

8.

```
> resid(model) #List of residuals
          1           2           3           4           5           6           7           8
 3.26143390 -3.00429730  4.80144705 -0.19000219 -1.55286121 -0.64143828 -0.19000219  0.80999781
          9          10          11          12          13          14          15          16
-9.74711686  3.35001096  0.35001096  1.80999781  0.08427976  8.53571585 -0.55286121 -2.27857926
         17          18          19          20          21          22          23          24
-2.27857926  1.35856172 -1.46428415 -1.01284806  0.71286998 -0.91572024  0.35856172  2.17285683
         25          26          27          28          29          30
 0.80999781  1.98715194 -0.55286121 -1.73001535 -3.01284806 -1.27857926
```



9.

10.

model performance has become better after the transfromation.